



Benton, M. J. (2015). Palaeodiversity and formation counts: Redundancy or bias? *Palaeontology*, 58(6), 1003-1029.
<https://doi.org/10.1111/pala.12191>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1111/pala.12191](https://doi.org/10.1111/pala.12191)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final open access version of the following article: Benton, M. J. (2015), Palaeodiversity and formation counts: redundancy or bias?. *Palaeontology*, 58: 1003–1029. Available at DOI: 10.1111/pala.12191

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

PALAEODIVERSITY AND FORMATION COUNTS: REDUNDANCY OR BIAS?

by MICHAEL J. BENTON

School of Earth Sciences, University of Bristol, Bristol, BS8 1RJ, UK; e-mail: mike.benton@bristol.ac.uk

Typescript received 2 April 2015; accepted in revised form 6 July 2015

Abstract: A key question in palaeontology is whether the fossil record taken at face value is adequate to represent true patterns of diversity through time. Some methods of assessing data quality have depended on the commonly observed covariation of palaeodiversity and fossiliferous formation counts through time, based on the assumption that the count of formations containing fossils, to a greater or lesser extent, drives diversity; but what if diversity drives formations? Close study of two fossil records, early tetrapods (Devonian–Jurassic) and dinosaurs, shows how the relationship between new taxa and new fossiliferous formations varies through research time. Initially, each new find represents a new fossiliferous formation and discovery follows the ‘bonanza’ model (fossils drive formations). In unexplored parts of the world, new taxa are identified frequently in new regions/formations. Only after time, in well-explored continents such as Europe and North America, does collecting style switch to a mix of exploration for new formations and re-sampling of known fossiliferous formations. Data are most striking for dinosaurs, where the Triassic–Jurassic record largely comprises finds from Europe and North America, where new formation discoveries reached their

half-life in 1914. This contrasts with the Cretaceous, which is dominated by rapidly rising discoveries from regions outside Europe and North America and the formation half-life for these ‘new’ lands is 1986, showing that 50% of new Cretaceous dinosaur-bearing formations were identified only in the past 30 years. The relationship between dinosaur-bearing formations and palaeodiversity then combines three signals in variable amounts, reflecting the original diversity (relative abundances of particular taxa in different formations), redundancy (new fossiliferous formations accruing because of new fossil finds) and sampling (intensity of exploration for new fossiliferous formations, and of search within already-sampled formations). For fossil vertebrates at least, formation counts of various kinds are poor predictors of sampling, missing, for example, the bonanza samples of Lagerstätten such as the Yixian Formation in China: thousands of specimens, dozens of species, but counted as one formation. These observations suggest that formation count cannot be regarded as an unbiased metric of sampling.

Key words: sampling, palaeodiversity, formations, quality of fossil record.

IN a succinct pair of articles in 1977, Peter Sheehan and David Raup laid out both sides of a key debate concerning the quality of the fossil record: was it ‘a reflection of labor by systematists’ (Sheehan 1977) or do ‘systematists follow the fossils’ (Raup 1977)? In other words, can we use cumulative sampling metrics (in this case, counts of publications) to account for sampling or are these metrics the outcome of vagaries in fossil occurrence? Sheehan argued that publications drive fossils, whereas Raup suggested that fossils drive publications, characterizing, respectively, the ‘bias hypothesis’ and the ‘redundancy hypothesis’.

A key problem in palaeobiology and macroevolution is to know how reliably the fossil record documents the history of the diversity of life. Without independent evidence, it is hard to determine whether palaeodiversity time series are dominated by biological signal or bias

(Smith and McGowan 2007; Benton *et al.* 2011). Bias and error in fossil data depend on four factors: rock volume, rock heterogeneity, accessibility and human effort (Raup 1972). At times, palaeontologists have veered from an optimistic view that the fossil record more or less documents the history of life (Sepkoski *et al.* 1981; Benton 1995; Stanley 2007) to a pessimistic view that much of the signal is error (Raup 1972; Peters and Foote 2002; Alroy 2010a). In either case, palaeontologists have recognized that they must seek to identify error and, where possible, correct for it.

Fossil record bias has commonly been estimated by the use of *sampling metrics*, or sampling proxies, time series taken from the rock and fossil records, or measures of human effort, that are said to document some aspects of bias (Smith 2001; Smith and McGowan 2007; Benton *et al.* 2011). Whilst sampling metrics can be

gathered independent of fossil data (e.g. rock area, total formation counts), it has been common practice to use the number of geological formations, number of localities or number of collections, compiled from the same literature that documents the fossil taxa as a proxy for sampling. These are, in order of inclusivity: (1) a *collection* – an assemblage of fossils from one location that were amassed in a single effort, or linked series of efforts, commonly documented in a single published paper; (2) a *locality* – a fossiliferous site that may be a natural exposure, such as a cliff or crag, or an artificial site such as a quarry or road cut; and (3) a *geological formation* – a named packet of rocks with limited and defined vertical and areal extent. Geological formations are generally mappable units, defined on the basis of outcrop, facies and contacts with formations below and above. Collections, localities and formations bearing the fossil taxa of interest have been regarded as useful sampling metrics, especially formation counts in one form or another (Peters and Foote 2001, 2002; Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011; Brocklehurst *et al.* 2012; Benson and

Upchurch 2013; Newham *et al.* 2014) because they incorporate aspects of all four biasing factors, rock volume, rock heterogeneity, accessibility and human effort. Additional sampling metrics are as follows: rock outcrop area (map area) for rocks of different ages, as a measure of rock volume (Raup 1976; Sepkoski 1976; Smith and McGowan 2007); and counts of published papers or active researchers, as a measure of human effort (Sheehan 1977; Alroy 2010b).

The strongest evidence for global-scale bias in the fossil record is the observation that palaeodiversity and sampling metric time series generally covary (Figs 1, 2). Indeed, the quality of agreement between the signals can sometimes be startlingly high. This covariation between fossil and rock record metrics has been explained in three ways: (1) the *bias model* says that the rock record drives the fossil record (Peters and Foote 2001; Smith 2001; Peters and Foote 2002; Smith 2007); (2) the *common cause model* says that both rock and fossil records are driven by a third factor, perhaps sea level change in the case of marine data (Peters 2005); and (3) the *redundancy model* says that the fossil

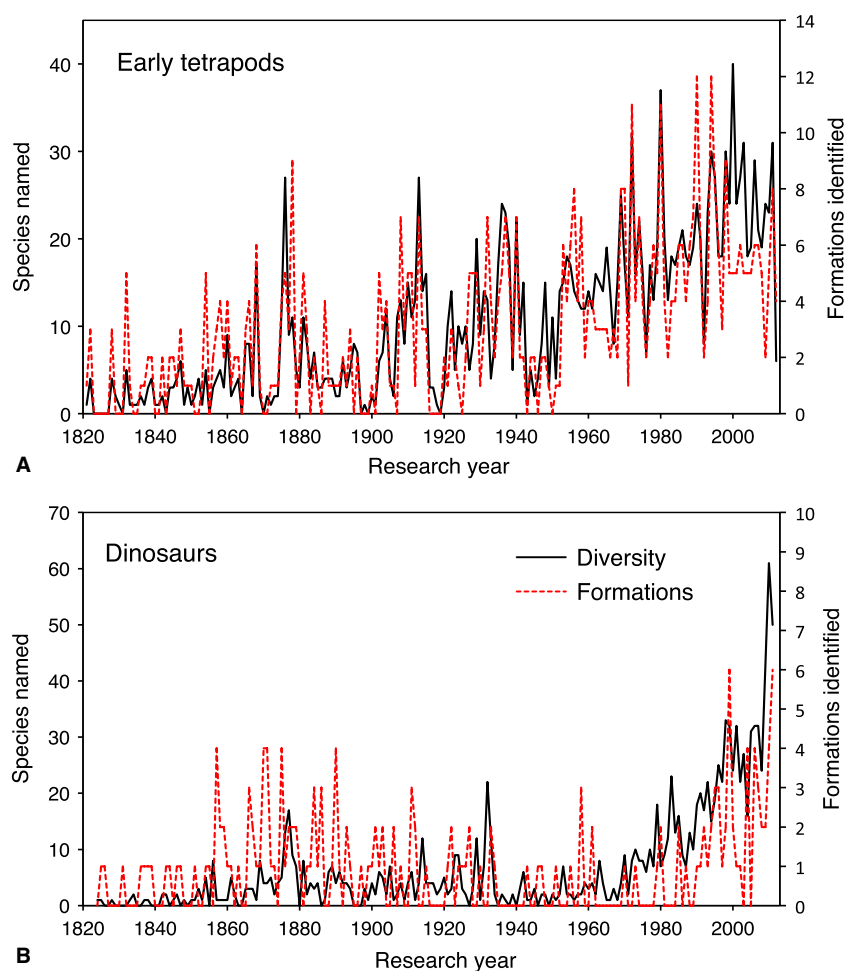


FIG. 1. Covariation of fossil-bearing formations and the fossils recorded from those formations through research time. Palaeodiversity time series (solid line) compared to formation counts (dashed line), for early tetrapods (A) and dinosaurs (B). Records are both counts per year rather than cumulative. Colour online.

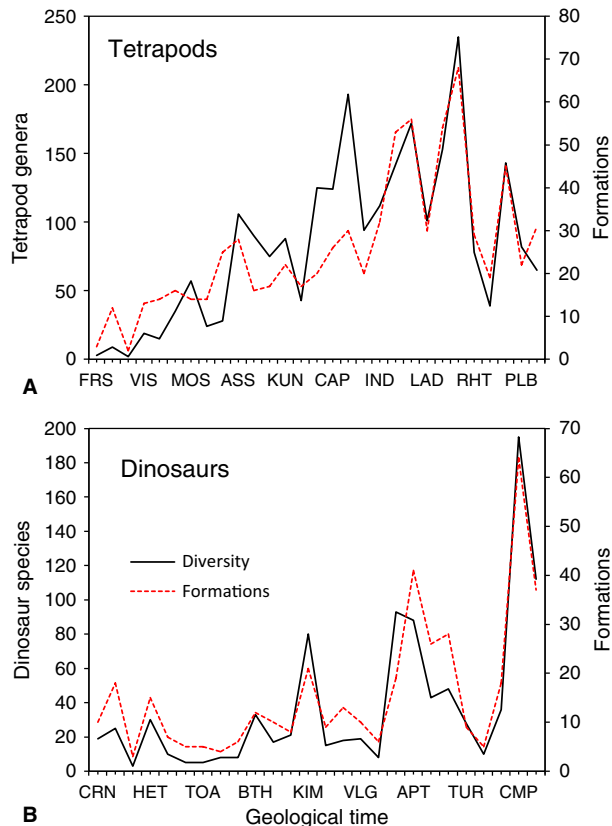


FIG. 2. Covariation of rock record and fossil record signals, through geological time. Palaeodiversity time series (solid line) compared to formation counts (dashed line), for early tetrapods (A) and dinosaurs (B). Abbreviations: APT, Aptian; ASS, Asselian; BTH, Bathonian; CAP, Capitanian; CMP, Campanian; CRN, Carnian; FRS, Frasnian; HET, Hettangian; IND, Induan; KIM, Kimmeridgian; KUN, Kungurian; LAD, Ladinian; MOS, Moscovian; PLB, Pliensbachian; RHT, Rhaetian; TOA, Toarcian; TUR, Turonian; VIS, Visean; VLG, Valanginian. Colour online.

record and sampling proxy signals may be partially redundant with each other (Benton *et al.* 2011). The redundancy may be empirical, in the sense that the two signals are the same or operational, in that we collect the data on both in non-independent ways. In simplified form, focusing on ‘what drives what’, the three viewpoints may be summarized as follows: bias (formations → fossils); common cause (environment → formations and fossils); redundancy (fossils ↔ formations, or fossils → formations). None of these three explanations is exclusive, and in reality, some components of all three hypotheses doubtless pertain in each case (Smith 2007; Benton *et al.* 2011; Hannisdal and Peters 2011). In the first studies to use statistical methods to detect directionality of signals, Hannisdal and Peters (2011) found that rock packages had bidirectional information transfer with palaeodiversity and, tellingly, Dunhill *et al.* (2014a) found that collections and formations

show bidirectional information transfer (= redundancy) with palaeodiversity.

Discriminating among these explanations for rock-fossil record covariation is difficult. Here, we focus on counts of formations that have historically yielded the taxa of interest, because these have been most widely used as a sampling metric in efforts to detect times of poor sampling and to provide corrected palaeodiversity time series. Formations have been used in two ways to assess sampling: as a source for a model of the effect of sampling on perceived extinction rates (Peters and Foote 2002) or as a source to model residuals as the portion of the palaeodiversity signal that cannot be explained by sampling (Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011; Lloyd 2012; Benson and Upchurch 2013; Lloyd and Friedman 2013; Newham *et al.* 2014). There are two assumptions behind the use of formations as a sampling proxy: (1) formation count is a metric of sampling, incorporating aspects of inadequacy of documentation relating to geology (non-preservation of rocks or fossils) and to human factors (variable sampling regimes); and (2) if strict formation count and palaeodiversity covary, then variability in the former *explains* variability in the latter (formations → apparent biodiversity).

The first claim has been widely assumed, or positively argued (Peters and Foote 2001; Smith 2001; Peters and Foote 2002; Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011; Lloyd 2012; Benson and Upchurch 2013; Newham *et al.* 2014). The second, directionality of the driver, has been less explored (Benton *et al.* 2011; Hannisdal and Peters 2011; Benton *et al.* 2013a; Dunhill *et al.* 2013, 2014a, b). What if palaeodiversity drives fossiliferous formation count in certain cases? For example, at a time of low global biodiversity, perhaps after a mass extinction event, global biodiversity will be low (Wignall and Benton 1999). Low biodiversity means few taxa, perhaps few fossils, and few counted fossiliferous formations. The claim is not that formation names are based on fossils, although this is often true – formations may be finely subdivided based on rock and fossil heterogeneity (Crampton *et al.* 2003; Smith 2007; Benton *et al.* 2011) – but simply that abundant fossils everywhere means many formations will be identified as fossiliferous and *vice versa*. Therefore, it is not a fact *a priori* that a low fossiliferous formation count means poor sampling. This study concerns only the relationships between the number of taxa and the number of formations containing those taxa through geological time. Several authors (e.g. Crampton *et al.* 2003; Smith 2007; Benton *et al.* 2011) have suggested that all collection-based cumulative metrics, including publication counts, author counts, locality counts and collection counts, harbour a considerable amount of redundancy with the palaeodiversity signal, and

yet some of these (most notably formation counts) have been widely used as sampling proxies.

Here, I assess the relationship between taxon counts and the numbers of formations that have yielded those taxa over time in two well-explored portions of the vertebrate fossil record: dinosaurs and early tetrapods. These are examples of 'sparse' fossil records, where sampling is known to be problematic, and where considerable attention has been devoted to exploring the covariation of rock record and fossil record. Plots of the discovery curves for new fossil taxa and new formations containing those fossils generally show close tracking through research time (Fig. 1). These case studies enable us to explore the ways in which knowledge has accumulated for these generally rare and poorly sampled taxa. In understanding the meaning of fossil-bearing formation counts (and perhaps related counts such as collections or localities) by following the palaeontologists into the field as they search for new fossils and new sources of fossils, perhaps a clearer understanding can be achieved of the relationship between the cumulative knowledge of fossils and the formations that yield them.

It is important to stress that this study does not consider the relationship between formation counts and the rock record, other than to argue that formation counts are a poor metric of rock volume or rock availability (Crampton *et al.* 2003; Smith 2007), and therefore to support the point that independent measures of the rock record are recommended for studies of fossil record quality (e.g. Crampton *et al.* 2003; Peters 2005, 2008; Smith 2007; Peters and Heim 2010; Hannisdal and Peters 2011; Dunhill *et al.* 2014a).

MATERIALS AND METHODS

Data

The data are taken from published sources. First, for the early tetrapods, the listing comprises 1388 genera, covering the first half of tetrapod history, from the Middle Devonian to Middle Jurassic (380–175 Ma), and was also compiled for earlier studies (Benton 2012; Benton *et al.* 2013a, b; Benton 2015, appendix 1 (ETD data)). This data set was expanded to species level, comprising 1959 species (561 amphibians, 779 reptiles, 619 synapsids; Benton 2015, appendix 2). We divide the early tetrapods into three subgroups, amphibians (paraphyletic, in traditional sense), reptiles (i.e. anapsids and diapsids) and synapsids. The second data set, for dinosaurs, was also compiled for earlier studies (Benton 2008a, b; Lloyd *et al.* 2008) and updated to the end of 2012. This provides a list of all 987 currently valid dinosaurian species, including Mesozoic birds (Benton 2015, appendices 1(Dino data), 3).

In both cases, data were compiled from the primary published literature. Species were listed, together with systematic, geological, geographic, taphonomic and biological data. Considerable effort was expended in documenting synonymies and other systematic corrections, which for dinosaurs reduced the total number of species ever named to less than 50% currently regarded as valid (Benton 2008a). This very high error rate is surprising, but taxonomic revisions must be taken into account before plotting a palaeodiversity time series to avoid random inflation of species counts. For the present study, particular attention was paid to the geological formation in which each tetrapod fossil occurred; these were documented and checked for synonyms (cases where a formation name used in the past has been revised, sometimes by subdivision). Further, the date of naming of each species was noted, so that the accumulation of knowledge through research time can be documented.

The date of 'discovery' of each fossil-bearing formation was also noted, as the date when the first fossil tetrapod or dinosaur was published from the formation (Benton 2015, appendix 4). A variety of possibilities exist for selecting a date of discovery for a dinosaur-bearing geological formation, namely: (1) the date of first discovery by any geologist, but not necessarily the date of naming; (2) the first naming of the rock unit in some sense, although formality in naming formations was not commonplace until after 1950; (3) the first discovery of a dinosaur; and (4) the first published discovery of a dinosaur. I chose the last of these options, mainly for reasons of practicality; to pursue any of the other three would require extended scholarly work on each formation, involving unpublished notebooks and other records; such data simply do not exist for the early work on geology in many parts of the world. Before 1850, such documentation was sporadic, but with the establishment of geological surveys after that time, most new geological work was probably published within 10 years after initial fieldwork, and so the variable lag time between definitions (1–3) and (4) might have reduced, or at least become somewhat standardized. Recording the date of discovery of a new dinosaur-bearing formation and of a new dinosaur species from the same publication guarantees to tie the two events together; however, this reflects the reality in many cases. An expedition goes out into new territory and finds a new dinosaur species in a formation that had not previously yielded dinosaurs all on a single day, and so the two discoveries are tied together. The point is that, in general, the 'formation count' is the 'fossiliferous formation count' and barren formations are excluded; therefore, seeking to document formations before target fossils were found would perhaps better document the sampling aspect, but this approach has not been used in the literature on quality of the vertebrate fossil record, although it

is correctly a core component of the use of independent macrostratigraphic metrics of rock volume (e.g. Peters and Foote 2001; Peters and Heim 2010).

In order to ensure some equivalence, formations were counted by name, whether formal or informal (e.g. 'Wealden', 'Continental Intercalaire'), but formal wherever possible. Full details of the formations-through-research-time are in Benton (2015, appendix 2). Decisions about which units are formations (and which are members or groups) follow published practice, although variants exist; for example, the Chinle and Dockum of the south-western United States are treated variously as groups or formations in the current literature. Here, I follow the current convention that both are groups, and document fossil occurrences by their constituent formations. These difficulties in scaling highlight the immense irregularity in the nature of formations as a metric of rock volume and rock availability – formations vary in volume over eight orders of magnitude (Benton *et al.* 2011), so they are clearly not comparable entities in terms of their claimed usefulness (e.g. Peters and Foote 2001) in documenting aspects of rock volume and rock availability. The difficulties of compilation illustrate difficulties in the use of formations as a sampling proxy, as noted before (e.g. Crompton *et al.* 2003; Smith 2007; Benton *et al.* 2011): their dimensions even depend on current politics, where many well-known formations cross state lines or country boundaries, for example from Montana to Alberta or from Mongolia to China, and yet they are given different formal names on either side of the border. I catalogue these as distinct formations. Although they are the same, geologically speaking, my aim is not to revise global geology, but to follow previous practice in such studies.

Here, I am using the 'strict formation count' – a version of the formation count that includes only those formations that have yielded fossils of the organisms in question, as was done earlier (e.g. Fröbisch 2008; Barrett *et al.* 2009). The aim here is to document discovery patterns of the new formations and new taxa to explore how the two might relate to each other. In subsequent studies, analysts have generally applied some version of a 'wider formation count' (Benton *et al.* 2011) that might be 'all fossiliferous formations' or 'all terrestrial formations', for example, so as to allow for non-occurrences, and so to some extent to explore whether drops in diversity might reflect a real drop or a failure in preservation or collection.

Throughout the study, all time divisions and their current best-estimate dates are taken from the latest international geological time scale (Gradstein *et al.* 2012).

It is important to stress that 'new formations' linked with 'new dinosaurs' does not mean that the formation was actually named in honour of the dinosaur – this

rarely, if ever happens (Benson and Upchurch, 2013) – and that has never been the key point about redundancy of formations and fossils. The point is that the roster of *fossiliferous* formations increases by one as each new dinosaur is discovered in a new basin or region; it is irrelevant why or how the formation was named. It is an entirely separate issue to consider the heterogeneity of the rocks that comprise formations, and the inevitable reduction in formation size when sedimentary heterogeneity is high and increase in formation size when heterogeneity is low. Heterogeneous and finely divided successions (such as the marine Jurassic of Europe) are formation-rich, and sometimes happen to be fossil-rich; there is a formations \leftrightarrow fossils link in these particular cases. This is not true for most vertebrate occurrences.

The taxon:formation ratio

The relationship between fossil finds and formation finds is explored by means of a taxon to formation ratio (T:F), simply the proportions of cumulative fossil finds and cumulative fossiliferous formations identified through research time. T:F ratios show different behaviours through research time, some remaining remarkably constant and others varying substantially. The T:F ratio wraps up a number of factors, some empirical (e.g. the original number of species per fauna/formation), some reflecting sampling (more formations provide more fossils) and some reflecting redundancy (if species are more geographically widespread at one time compared with another, the numbers of fossil-bearing formations are likely to be higher, all other things being equal). Exploring collecting histories and the T:F ratio may shed some light on the relative balance of these different components of the relationship between taxa and formations.

The simplest sense of redundancy of the T:F metric is that the discovery of a new dinosaur and of a new dinosaur-bearing formation necessarily occur at the same moment, and a steady T:F ratio could reflect that. The value of a redundant T:F relationship need not be a value of 1.0, but could be any other value that reflects the original mean numbers of species per formation/fauna. So, if there are typically five dinosaur species per formation, the T:F ratio could remain constant at a value of 5, reflecting the original, empirical faunal composition. Sampling bias, as expressed by the T:F ratio, can occur in two ways: (1) if a particular formation yields more or less than the average because of preservation issues or levels of effort by palaeontologists; and (2) in terms of the presence or absence of certain formations, representing a temporal-spatial sample of regional or global dinosaurian diversity.

The T:F ratio through research time may remain constant, or it may increase or decrease. In interpreting

changes of slope, it is assumed that the running tallies of new fossiliferous formations and new taxa cannot decline – this is because we use current opinion in recording discoveries of formations and taxa, so retrospective synonymy or correction of other errors cannot be a reason for decline. If the T:F value remains constant, this indicates that new species and new fossiliferous formations are being identified at a constant rate. The mean number of species per formation (fauna) is not changing because re-exploration of known formations, with perhaps rising species counts, is matched by the discovery of new fossiliferous formations with lower-than-average species counts. Sampling is clearly improving on a global scale, as more fossiliferous formations and more species are added to the rosters, but it cannot be said from a constant T:F value through research time whether sampling per formation is improving or not. If the T:F ratio rises through research time, this indicates that the number of new taxa is increasing faster than the number of new formations, as a result of intensified efforts in the field collecting new specimens, intensified efforts in the museum identifying new species from old specimens, or simply that the new formations just happen to be richer on average than those already identified. Analogously, if the T:F ratio falls through research time, this indicates that the number of new taxa is increasing more slowly than the number of new formations, either because of less effort being expended in the field and museum, leading to fewer finds per formation (poor sampling), or because the new formations just happen to be less rich on average than those already identified, either because the faunal diversity was lower or fewer fossils are preserved (empirical).

Time series comparisons

Rock and fossil time series were compared initially by plotting both through research time (Fig. 1) and through geological time (Fig. 2). In the latter case, each species and each geological formation is counted once, marking the first discovery of the species in question, and assigned to the relevant stage-level division of geological time. If a species was subsequently found in a different formation that had not previously yielded fossils of the group in question, this occurrence was added to the ‘new formations’ count, but not to the ‘new species’ count. Among tetrapods, most multiple occurrences are within the same time bin (geological stage or 11 myr bin) and so this does not affect comparisons of our discovery data with palaeodiversity time series. Several early tetrapod and dinosaur genera span two geological stages, but only minute numbers of species occur in more than one time bin. Comparisons between time series were assessed by various standard correlation methods, including Pearson product-moment, Kendall tau and

Spearman’s Rank correlation, suitable for such nonparametric correlation problems. Further, the calculations were also carried out after generalized differencing of the time series, an accepted way to detrend the data and focus on shorter-term fluctuations apart from any long-term trend. Graeme Lloyd’s R code was used (<http://www.graemetlloyd.com/methgd.html>).

The second set of analyses consists of comparisons of discovery curves for taxa and formations. Collector (= discovery) curves were first used (Cain 1938) as a means of estimating the completeness of local or regional species lists: with continuing collecting effort, the number of new species reported rises fast, and then reaches an asymptote after a certain number of specimens have been collected, or after a number of days of search. In addition to their use as a tool for ecological and biodiversity sampling, discovery curves have been used widely to estimate global completeness of taxon counts, whether for extant groups (Bebber *et al.* 2007; Costello *et al.* 2012; Nabout *et al.* 2013) or for extinct groups (Benton 1998, 2008a; Purnell and Donoghue 2005; Tarver *et al.* 2007; Bernard *et al.* 2010; Brocklehurst and Fröbisch 2014). In all cases, the accumulation of species is considered globally, although comparisons between discovery curves for different continents may show, for example, that Europe and North America have been more thoroughly collected than other continents (Benton 2008a; Bernard *et al.* 2010). The shapes of discovery curves cannot be compared directly, so we use the half-life as an indicative, comparative measure. This discovery curve half-life was introduced by Bernard *et al.* (2010) as a means of documenting whether the rate of discovery had slowed or increased, and for comparing between clades or between taxon and formation discovery curves. In all cases, discovery curves are based exclusively on ‘now-valid’ taxa and formations, rather than ‘then-valid’ data; this ensures a single, current standard for inclusion of taxa. Our aim here is not to recreate a picture of what was believed to be true in 1932, including all the oddities of decisions at the time about validity of taxa and formations, but to remove as many confounding variables as possible so as to focus on the raw accumulation of knowledge from a single standpoint (the present day). To model contemporary reality would give false signals, such as a remarkable rate of discovery of dozens of new dinosaurs in 1932 that all stem from a single monograph and have since been shown to have been illusory. The ecologist hopes to report only valid new sightings when compiling a discovery curve, and we follow common practice (e.g. Alroy 2002; Benton 2008a; Bernard *et al.* 2010).

The third analysis is to consider whether the ratio of number of species and number of formations yielding those species (T:F ratio) changes through research time. An underlying assumption of much previous work in

which number of formations was used as a sampling metric (Peters and Foote 2001, 2002; Fröbisch 2008; Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011; Benson and Upchurch 2013) was that ‘formation count’ was the yardstick for comparison. It is simple, using the data series at hand to calculate a rolling T:F ratio through research time, from the discovery of the first fossil reptiles in the 1820s to the present day. The count of taxa comprises species of the clade in question, and ‘formations’ is the count of formations that have first yielded these species. The aim is to explore how the two time series, and the T:F ratio, vary through research time and to see whether a narrative approach might tease apart whether one drives the other, or whether both are intimately linked in a redundant relationship.

The fourth analysis is to compare different time bins of the taxa, namely Permian vs Triassic tetrapods and Triassic–Jurassic vs Cretaceous dinosaurs, to determine whether there are any differences, and if so, why. The null expectation is that there should be no differences between discovery data on taxa and the formations that yield them for different time-period partitions of the early tetrapods and the dinosaurs data. If there are differences, this could illustrate aspects of the interplay of the discovery of new fossiliferous formations and new taxa in new territories with the discovery of new taxa in known formations, or differences in sampling, based on geology or human behaviour, between the time partitions.

Distinguishing well- and less-well-sampled areas

It may be hard to identify a single, categorical test of the influence of sampling on any large-scale diversity-through-time data set, but comparisons between geographical regions where sampling began at different times in research history may help. For example, palaeontologists began collecting and identifying early tetrapod fossils and dinosaurs in Europe in the 1820s, in North America in the 1850s, in China in the 1920s and in other parts of Asia, Africa, Australasia and South America at various times between. Therefore, it might be predicted that a comparison of Europe + North America vs the rest of the world could illuminate something about the progress of the sampling of formations and taxa. This intuition is borne out by collector curves of early tetrapods (Bernard *et al.* 2010) and dinosaurs (Benton 2008a), which show that the rate of discovery of new taxa has slowed down in Europe and North America, whereas discovery rates continue to rise in some other continents, such as Asia and South America. It is widely understood that Europe and North America are better sampled palaeontologically than most other continents: Allison and Briggs (1993) noted this was the case for marine fossil Lagerstätten, Smith

(2001) for marine invertebrates in general, Fastovsky *et al.* (2004) for dinosaurs and Brocklehurst *et al.* (2012) for Mesozoic birds.

Therefore, we compare discovery histories for both early tetrapods and dinosaurs in ‘well-sampled’ vs ‘less-well-sampled’ continents, namely Europe + North America vs the rest of the world. Reflecting a historical quirk, we also plot a ‘rest of world, excluding South Africa’ set of curves, because the Karoo Basin in South Africa yielded tetrapod fossils very early, in the 1840s, and has been very actively hunted and documented ever since, so it might be thought to behave like an old, well-sampled region, more in line with Europe or North America than the rest of Africa.

RESULTS

Time series comparisons

The covariation between fossil and rock record time series has often been mentioned (e.g. Peters and Foote 2001, 2002; Fröbisch 2008; Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Benton *et al.* 2011; Mannion *et al.* 2011), and the records of early tetrapods and dinosaurs indeed show a relationship between species numbers and counts of formations yielding those species, whether plotted against research time (Fig. 1) or geological time (Fig. 2). In the plots against research time (Fig. 1), there is no substantial trend through time, except for a rise of both signals since 1970. Peaks and troughs in the raw data are highly significantly correlated throughout, for both early tetrapods (Fig. 1A) and for dinosaurs (Fig. 1B).

The strong covariation between rock and fossil records through geological time, for both early tetrapods (Fig. 2A) and dinosaurs (Fig. 2B), is also evident. At times, the peaks in both curves appear to be out of synchrony, for example during the Carboniferous and Early Permian for early tetrapods and during the Early Cretaceous for dinosaurs (diversity peaks in the Barremian–Aptian, formations peaks in the Aptian–Cenomanian). After generalized differencing, all correlations remain highly significant (Table 1), confirming that part of the correlation arises from the overall rising trend in values through geological time, but the details of fluctuations also show close correspondence in both pairs of time series.

Formation discovery and taxon discovery

Discovery curves for taxon counts and the formations that yield them through research time are closely similar

TABLE 1. Time series comparisons of cumulative formation counts and new taxa (species, genera) counts.

Analysis	Pearson's corr. coeff.	p-value	Spearman's rho	p-value	Kendall's tau	p-value
Tetrapods	0.85178	4.636×10^{-9}	0.80631	2.1951×10^{-9}	0.70650	9.532×10^{-8}
Dinosaurs	0.93630	2.122×10^{-12}	0.94033	9.897×10^{-13}	0.81576	9.911×10^{-9}
GD tetrapods	0.77710	1.153×10^{-6}	0.62616	0.00048	0.45503	0.00050
GD dinosaurs	0.89612	1.397×10^{-9}	0.86077	2.208×10^{-6}	0.72667	1.453×10^{-8}

Data analyses are given for raw data (top two rows) and for generalized-differenced (GD) data (second two rows). All three standard correlation methods are shown: Pearson product-moment correlation (parametric) and Spearman's rank correlation and the Kendall rank correlation (both nonparametric), with two-sided p-values. All p-values indicate very highly significant correlations ($p < 0.001$).

for all six data sets (amphibians, reptiles, synapsids, amniotes, tetrapods, dinosaurs; Fig. 3). All of these show that the cumulative formations discovery curve is slightly ahead of (= is less concave than) the taxon discovery curve, except for synapsids (Fig. 3C) in which the taxon discovery curve crosses over the formations discovery curve around 1900–1920, suggesting that for this clade at least more new taxa were being discovered in established formations than for the other clades. Comparing such curves is difficult as they are plotted on differing y-axes (maxima aligned at top of y-axes) and the T:F ratio is more informative; this is highest for Synapsida of all six clades (Table 2). This difference may be a result of the scaling of formations – the Karoo Permo-Triassic formations and assemblage zones, from which many of the synapsid taxa have been identified, are huge, each

equivalent to a sizable span of time and an enormous geographical area, often with excellent exposure, so the chances are high of finding many new taxa in each formation, but the effect of areal extent of formations needs testing.

All six discovery curves (Fig. 3) show similar, hollow shapes, with no sign of an asymptote, suggesting that new taxa and the formations that yield them continue to be discovered at an increasing rate for all clades. The rate of discovery varies between clades (Table 2), and the curve shape can be determined by comparing their half-lives (the year in which they reached half their current estimated diversities (Bernard *et al.* 2010)). Half-lives for taxon discovery range from 1954 (Synapsida) to 1977 (Reptilia) and for formation discovery, from 1953 (Amphibia) to 1972 (Reptilia). The Dinosauria stands at an extreme, with

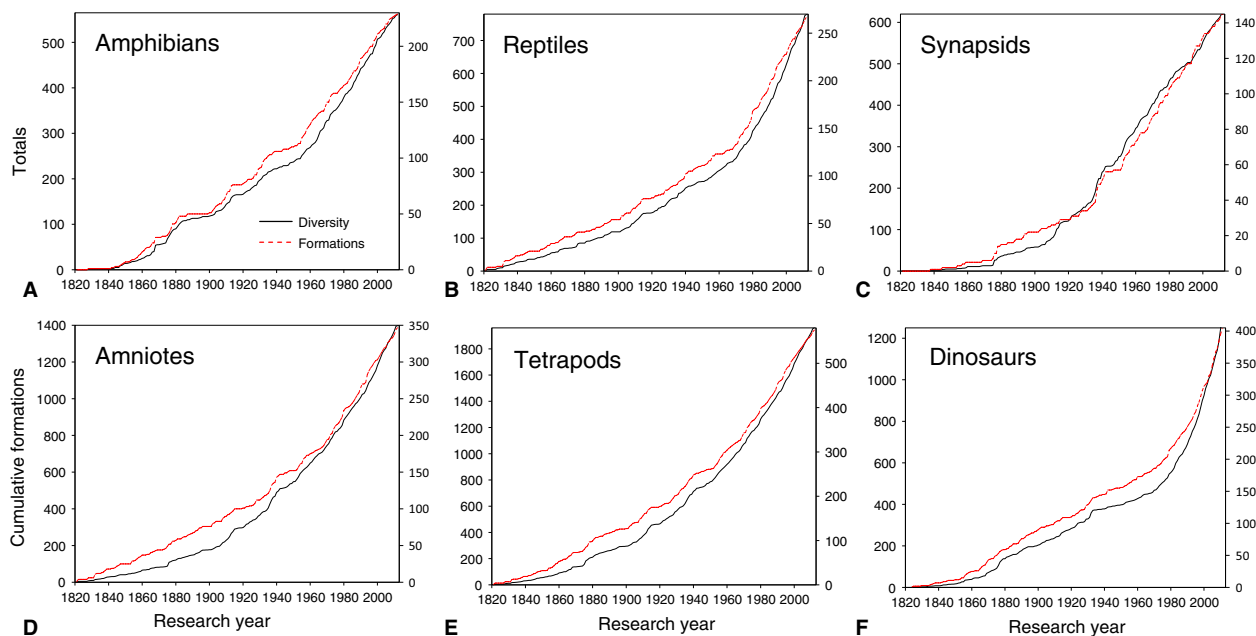


FIG. 3. Cumulative discovery of taxa and fossiliferous formations for early tetrapods and dinosaurs. Discovery curves for five divisions of early tetrapods, from Devonian to Early Jurassic (A–E), and dinosaurs (F), showing cumulative curves for species or generic diversity (solid line) and counts of fossiliferous formations (dashed line). Colour online.

TABLE 2. Timing of discovery of new taxa and new fossiliferous formations for six tetrapod clades, from 1821 to 2012.

Clade	Total taxa	Total formations	2012 T:F	Half-life taxa	Half-life formations
Amphibia	561	229	2.45	1963	1953
Reptilia	779	267	2.92	1977	1972
Synapsida	619	145	4.27	1954	1960
Amniota	1398	347	4.03	1965	1962
Tetrapoda	1959	501	3.91	1965	1957
Dinosauria	1294	402	3.22	1987	1977

For the first five, the taxa are genera, for dinosaurs the taxa are species. Note that Amniota comprises Reptilia + Synapsida; Tetrapoda comprises Amphibia + Amniota. Taxon totals add up, but formation totals do not because many formations yield examples of several major clades. The half-life is the date at which half the current (2012) total had accumulated.

half-lives for both taxa and formations occurring at much more recent dates (1987 and 1977 respectively), presumably an indicator of the intense recent and current interest in discovering new dinosaurs in new locations.

When cumulative totals of formations and species are compared, they show straight-line relationships (Fig. 4), for which the correlation coefficients are very high

($R^2 > 0.99$ in all cases) and the probability values show very highly significant correlation ($p \ll 0.001$). This suggests that for any of the tetrapod data sets, however they are divided, the discovery of new taxa and the discovery of new formations yielding those taxa have been very tightly linked through research time.

When we track the six examples through research time (Fig. 5), the T:F ratios all began low and fluctuated substantially in the years from 1820 to 1860, presumably as a result of small sample sizes. After that, the T:F ratios rose according to various patterns, some apparently stabilizing after 1870 (amphibians, dinosaurs), 1900 (synapsids) or 1950 (amniotes), although all with a rise in the past 20 years, and others (reptiles, tetrapods) rising more or less steadily towards the present day. The current T:F ratios (Table 2) show substantial differences between groups, ranging from 2.45 (amphibians) to 4.27 (synapsids), but these differing levels appear to have been steady, or at least distinct, for each clade for a century or more. The differing T:F ratios could indicate either that amphibians have always been relatively low-diversity components of ecosystems, whereas synapsids have been twice as diverse, or that synapsids are preserved and found twice as commonly as amphibians. Variations in the T:F ratio do not simply reflect sample size: values are lower for all 1959 species of

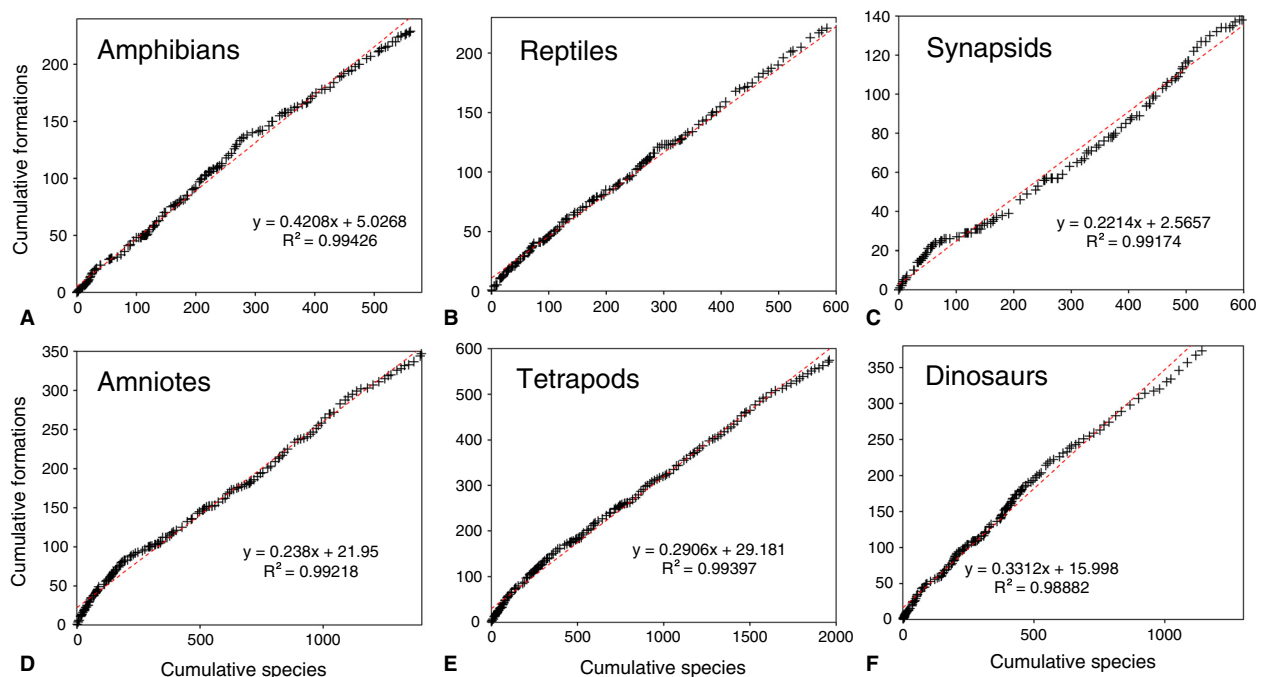


FIG. 4. Cumulative discovery of taxa and fossiliferous formations for early tetrapods and dinosaurs. Plots for five divisions of early tetrapods, from Devonian to Early Jurassic (A–E), and dinosaurs (F). In all cases, the correlations between cumulative formations and cumulative taxa are very highly significant ($p \ll 0.001$), with highest correlation for reptiles in general and lowest for dinosaurs, where formation discovery exceeds taxon discovery in the middle of the curve (mid-twentieth century), and taxon discovery exceeds formation discovery to the right, representing the last 20 years. These curves suggest that palaeontologists are still in a very early stage of knowledge of these clades, in which new fossiliferous formations continue to be discovered, and these typically yield new taxa. Colour online.

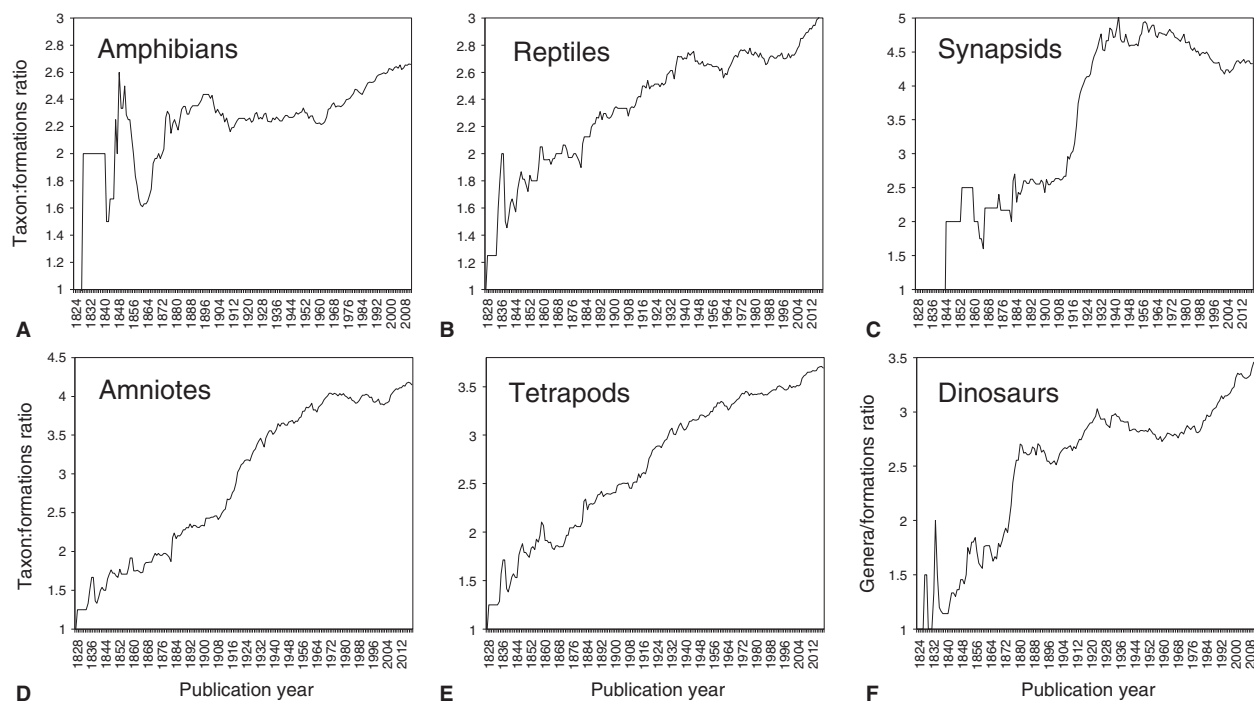


FIG. 5. Taxon:formation ratios for early tetrapods and dinosaurs. Plots of the ratios for five divisions of early tetrapods, from Devonian to Early Jurassic (A–E), and for dinosaurs (F), express the relationships between the coupled discovery curves of taxa and formations. They show very different patterns, some stabilizing after 1870 (amphibians, dinosaurs), 1900 (synapsids), 1950 (amniotes), and others (reptiles, tetrapods) rising more or less steadily towards the present day. The T:F ratio shows changing relationships between discovery of new formations and new taxa.

tetrapods (3.91) than for all 1398 species of amniotes (4.03). This is because the 561 species of amphibians show lower T:F ratios, bringing the overall value down.

The upturn in T:F ratios for reptiles and dinosaurs (Fig. 5B, F) may suggest that most finds in the past 50 years of research have been from formations already known to yield reptiles/dinosaurs, and so these finds represent an improvement in sampling of those formations. The recent drop in T:F ratio for synapsids (Fig. 5C), on the other hand, suggests that a number of finds have been made in previously unfossiliferous formations or previously geologically unexplored parts of the world, but that the number of taxa recorded per formation is lower in these newly discovered sources than in those previously identified. For example, if sampling of those newly identified formations is poorer than for those that have been known longer.

Comparing collecting histories

Permian and Triassic tetrapods. Discovery curves and plots of T:F ratios for Permian and Triassic tetrapods (Fig. 6) show differences. The sum totals of species identified are similar (661 Permian, 890 Triassic), and the

‘formations-led’ discovery curves are also similar. However, the discovery curves for Permian tetrapods (Fig. 6A) are less concave than those for Triassic tetrapods (Fig. 6B) and their half-lives differ accordingly (1951 for Permian tetrapods, 1975 for Triassic tetrapods; Table 3). The T:F ratio plots (Fig. 6C, D) show that values for Permian tetrapods have remained constant at about 4.5 since 1930, whereas Triassic values continue to rise towards 3.8. There are two phenomena to explain here, namely the differing mean T:F values between Permian and Triassic, and the differing collector curves. In summary, the first may be broadly real and the second could reflect differing sampling practice.

First, the differing mean T:F ratios could say something about the original diversities and distributions of faunas, or about preservation probability, collecting intensity, differing taxonomic practice and the balance of well-sampled (i.e. Europe and North America) and poorly sampled (rest of world) continents. These last four factors, all of which are aspects of sampling, are considered in turn. In the key Permo-Triassic basins, such as the Karoo, Russia and parts of the south-western United States, the rock types in the Permian and Triassic are similar, topography and exposure as a result of weathering and erosion are similar, and field crews hunt for fossils in similar ways. Collecting

FIG. 6. Comparison of discovery curves and taxon:formation ratios. Discovery curves (A–B) and T:F ratios (C–D), for Permian (A, C) and Triassic (B, D) tetrapods, show differing patterns between the two time periods, with steady accumulation of both new species and new fossiliferous formations through the Permian, rendered as a stabilizing T:F ratio, after 1940, and constantly accelerating discovery curves for the Triassic, shown as a continually rising T:F ratio. These differences, and especially the absolutely different numerical values for the T:F ratios for Permian and Triassic, provide information about both the original faunal compositions and about sampling. Colour online.

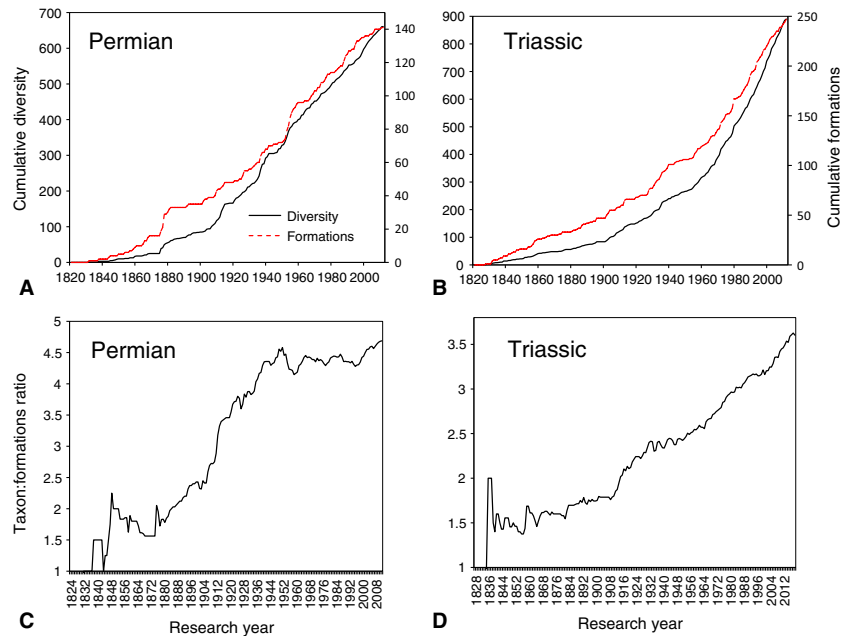


TABLE 3. Timing of discovery of new taxa and new fossiliferous formations for divisions among early tetrapods and for dinosaurs.

Sample	Total taxa	Total formations	2012 T:F	Half-life taxa	Half-life formations
Perm. tetrapods	661	136	4.86	1951	1948
Tri. tetrapods	890	235	3.79	1975	1965
Tri.–Jur. dinosaurs	277	102	2.72	1986	1942
Cret. dinosaurs	988	272	3.63	1988	1984
Tetrapoda – old	1168	402	2.91	1964	1954
Tetrapoda – new1	792	130	6.09	1966	1969
Tetrapoda – new2	464	134	3.46	1978	1970
Dinosauria – old	602	157	3.83	1953	1914
Dinosauria – new	694	218	3.18	1994	1986

The data are partitioned into Permian and Triassic tetrapod species, and Triassic–Jurassic and Cretaceous dinosaur genera, from 1821–2012. In addition, tetrapods are divided into those sampled from Europe and North America (old), the rest of the world (new1) and the rest of the world excluding South Africa (new2). In addition, similar comparisons of dinosaurs from well-sampled areas (Europe and North America; ‘old’) and less-well-sampled areas (rest of world; ‘new’) are listed. The half-life is the date at which half the current (2012) total had accumulated. *Abbreviations:* Cret., Cretaceous; Jur., Jurassic; Perm., Permian; Tri., Triassic.

intensity ought to be similar because of the noted geological factors, and these control access and visibility of the rock. Frequently, the same field crews work across Permian and Triassic redbed successions in restricted geographical areas, sometimes deliberately working across the Permo-Triassic boundary in their attempts to document the pre- and post-extinction faunas. Further, the same palaeontologists in many cases describe the fossils from above and below the boundary, tracking their chosen taxonomic group and applying the same standards in description and species identification. Clearly, there must be differences in all these sampling factors further from the Permo-Triassic boundary, between, for example, the

rich early Permian redbeds of the south-western United States and the equally productive Late Triassic redbeds of Argentina and South Africa. Finally, in terms of the balance of sampling from ‘old’ and ‘new’ lands, for Permian tetrapods, 43 of the 141 formations (30%) are from ‘new’ lands; for Triassic tetrapods, 91 of the 247 formations (37%) are from ‘new’ lands. These figures are similar, and perhaps contrary to expectations if this were the explanation: formations from ‘new’ lands, especially from South Africa and China, are often very extensive, which would lead to higher T:F ratios if this were a contributory factor; in fact, the opposite is the case, with Permian T:F ratios higher than Triassic T:F ratios (Fig. 6).

Perhaps then, much of the persistent difference in T:F ratios between the Permian and Triassic tetrapod collecting records could reflect underlying reality in one of two ways: either increasing diversity per formation, or increasing endemism. For the first, it might be that Permian formations contain more diverse tetrapod faunas than Triassic formations, and the low Triassic figures could reflect the devastation of the Permo-Triassic mass extinction and the slow rebuilding of faunas in the Triassic, most of which, during the Early, Middle and early Late Triassic were under-strength. In addition, palaeobiogeographical evidence suggests that Triassic faunas were less cosmopolitan and more endemic than in the Permian (Sahney and Benton 2008; Sidor *et al.* 2013). If faunas are cosmopolitan, T:F will remain constant or decline with increased sampling, whilst with endemic faunas, it will increase.

Thus, the continuing rise in the Triassic T:F ratio through research time (Fig. 6D) may reflect equal sampling of more endemic faunas, combined with massively increased interest in Triassic tetrapods in the past 30 years, perhaps with a focus on the origin of dinosaurs, in other words intensification of sampling of largely known formations, so driving up the ratio of taxa per formation. This recent intensification of interest is con-

firmed by the discovery curve half-lives (Table 3), which are much more recent for the Triassic (1975, 1965) than for the Permian (1951, 1948). However, note that the half-life for Triassic formations (1965) is rather earlier than for taxa (1975), confirming that many new taxa reported after 1975 come from already identified tetrapod-bearing fossiliferous formations.

These differences in corresponding taxon and taxon-bearing formation counts for Permian and for Triassic tetrapods, as expressed by their substantially different half-lives, allow us to reject the null hypothesis that counts of fossil-bearing formations predict taxon counts exactly. Something other than simply the sampling of formations yielding tetrapods is needed to explain why recorded diversity has changed between the Permian and Triassic. We can reject the idea of redundancy (one new fossil, one new formation with that fossil) in this case. Whether this is biological in origin or reflects some other aspect of artefact needs further exploration.

Triassic–Jurassic and Cretaceous dinosaurs. The comparison of discovery curves and T:F ratios for dinosaurs of the Triassic and Jurassic (combined because of low totals for the Triassic alone), and of the Cretaceous, also shows surprising differences (Fig. 7). Both are, again, ‘forma-

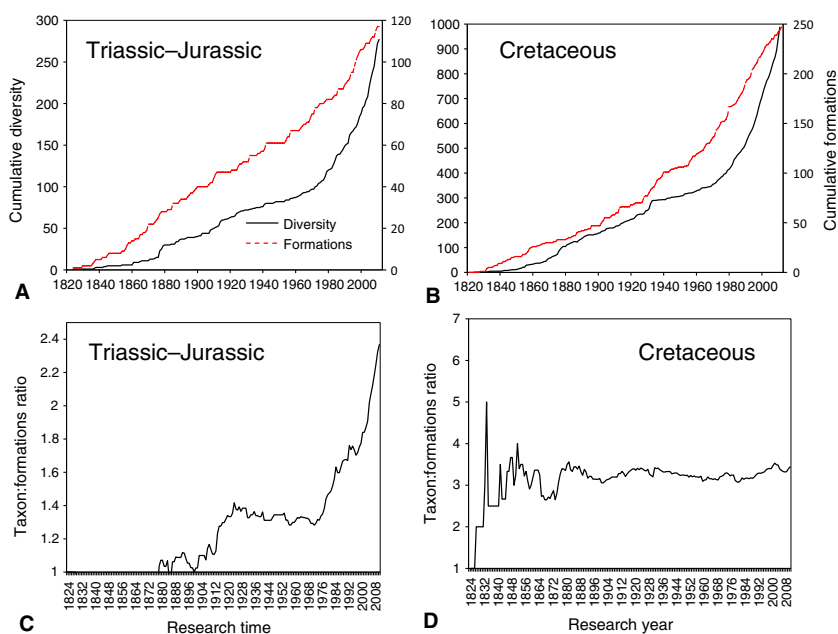


FIG. 7. Comparison of discovery curves and taxon:formation ratios. Discovery curves (A–B) and taxon:formation ratios (C–D), for Triassic–Jurassic (A, C) and Cretaceous (B, D) dinosaurs, show differing patterns between the two time periods. There is a steady accumulation of new fossiliferous formations, and accelerating recovery of new genera, through the Triassic–Jurassic, rendered as a constantly accelerating discovery curve and a continually rising T:F ratio. On the other hand, the Cretaceous shows more parallel discovery curves, and the T:F ratio, remarkably, has been constant since 1850. These differences, and especially the absolutely different numerical values for the T:F ratios for Triassic–Jurassic and Cretaceous, provide information about both the original faunal compositions and about sampling. Colour online.

tion-led', like most of the large subclade samples (Fig. 3). The Triassic–Jurassic dinosaur sample (Fig. 7A) is characterized by the current total of 277 dinosaur genera produced from a relatively small number of formations; half of these dinosaur-bearing formations had been identified by 1942 (Table 3). For the Cretaceous on the other hand (Fig. 7B), there are many more dinosaurs (988 genera) and the formations half-life date is 1984, indicating continuing high rates of discovery of dinosaur-bearing formations in the past 30 years. The taxon discovery curves are much more similar between the Triassic–Jurassic and Cretaceous samples, however, with half-lives in 1986 and 1988, respectively (Table 3), reflecting their shared steepening after 1980, especially in line with the influx of many new taxa from China. These very different formation discovery tracks are reflected in the plots of T:F ratios (Fig. 7C–D), with continuing increases in values for the Triassic–Jurassic, but an extraordinary steady state in the T:F ratio for Cretaceous dinosaurs since 1850 or 1880.

The fact that discovery patterns differ between the Triassic–Jurassic and Cretaceous dinosaur samples (Fig. 7) is not unexpected, having been noted before (Benson and Mannion 2012). The discovery curves for taxa (Fig. 7A–B) are similar in shape and confirm the null expectation that dinosaurs are equally sought and equally easy (or difficult) to find for all broad-scale divisions of the Mesozoic. The differences in T:F ratios arise almost entirely from differences in the dinosaur-bearing formation discovery curves: whereas that for the Triassic–Jurassic is nearly straight, indicating a remarkably steady rate of accumulation through research time, that for the Cretaceous is more concave and more closely coupled with the taxon discovery curve (and more similar to the wider sample of case studies; Fig. 3).

Perhaps the differences in discovery patterns between Triassic–Jurassic and Cretaceous dinosaurs are largely geological, reflecting two issues: the relative quantities of rock available, and their distribution between regions. First, there is much more Cretaceous rock outcrop worldwide than rock of Triassic or Jurassic age (Wall *et al.* 2009, fig. 3); this is true for both marine and continental sediments. This presumably correlates with the fact there are more than twice as many Cretaceous dinosaur-bearing formations as for the Triassic and Jurassic combined, and hence a greater opportunity to discover new formations with dinosaurs in the Cretaceous. This greater outcrop area and greater number of unique dinosaur-bearing formations in the Cretaceous would increase the global diversity of dinosaurs in the Cretaceous by offering improved sampling worldwide.

The second geological factor is the distribution of those Cretaceous rocks and dinosaur-bearing formations. In fact, the Cretaceous is dominated by dinosaur-bearing

formations from outside the European and North American continents, comprising 164 of 272 formations (60%), but only 48 of 102 formations (47%) for the Triassic–Jurassic. Hence, collecting worldwide in the past 20 years has had a substantially different character between the Triassic–Jurassic on the one hand, where palaeontologists have been devoting their efforts to re-exploring known dinosaur-bearing formations, whether by making new collections in the field or re-studying museum collections from which they name new species, and the Cretaceous on the other, where new formations, especially outside Europe and North America, are contributing massively to the discovery of new dinosaurs.

None of these geological factors (more rock, more formations and more 'new lands' to explore in the Cretaceous), however, can directly explain the differing T:F ratios. The higher T:F value for the Cretaceous than for the Triassic–Jurassic could then represent something about the original faunas. There may just have been more dinosaurian species per formation in the Cretaceous than in the Triassic and Jurassic (see below).

Distinguishing well- and less-well-sampled areas

The historical plot (Fig. 8) confirms the qualitative intuition (see Material and Methods) that for both early tetrapods and dinosaurs, Europe and North America dominated early collecting and have been sampled for the longest period, and perhaps also more thoroughly than other continents. For early tetrapods (Fig. 8A), 'rest of world' samples emerged relatively early, in the 1840s, with the first finds in South Africa and Asiatic Russia, and the proportions of Europe–North America vs 'rest of world' formation counts varies episodically, decade-by-decade. Perhaps surprisingly, the contribution of fossiliferous formations for early tetrapods from 'newer' continents never exceeded 50% and remained at 20–50% throughout the twentieth century. In other words, for early tetrapods, there are still, on average, more new fossiliferous formations being discovered in Europe and North America combined than in the rest of the world together.

The story is very different for dinosaurs (Fig. 8B), perhaps reflecting the overall much more intensive efforts being expended by much larger armies of palaeontologists everywhere in the world, and that there were many more dinosaurs in the larger area of the world outside Europe and North America, especially in the well-documented Cretaceous. The pattern remains somewhat similar to that for early tetrapods up to 1890 or 1900, but through the twentieth century it is very different. The contribution made by newly discovered dinosaur-bearing formations in the rest of the world reached 80% in the 1920s and has varied between 65% and 85% since then. In other words,

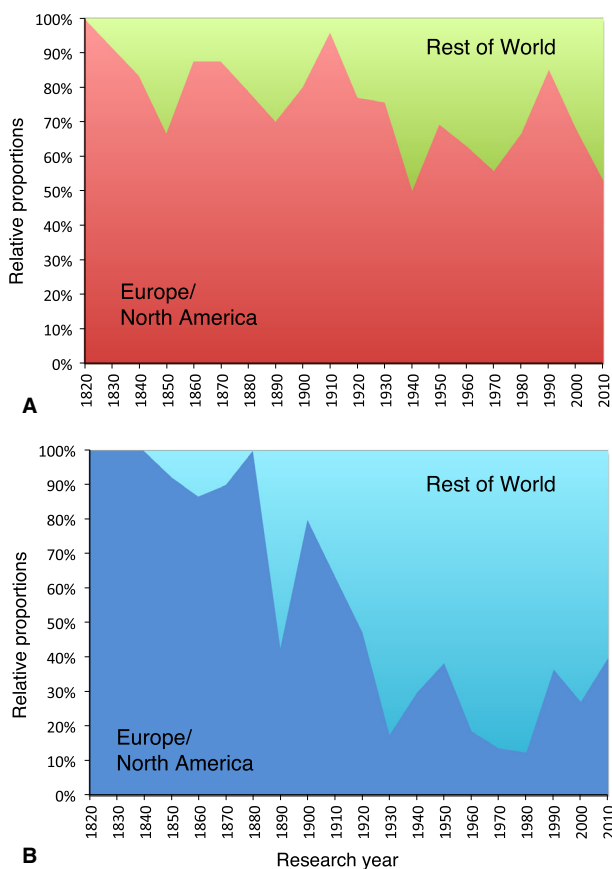


FIG. 8. Historical documentation of sampling of early tetrapods (A) and dinosaurs (B), showing how the traditional research areas (i.e. Europe and North America) were nearly the sole sources of fossils up to 1850, but how the ‘rest of the world’ took a greater and greater share through research time, reaching 40% for early tetrapods and 70% for dinosaurs. Data are the sum totals of new formations identified for each broad geographical realm, binned by decades. Colour online.

some new dinosaur-bearing formations continue to be identified in Europe and North America, but the vast majority for the last 100 years have come from China, South America and other parts of Asia, Africa and Australasia.

Analysis of the patterns behind this narrative is fascinating. For early tetrapods, the cumulative collecting curves are very different between all three subsets of the data (Fig. 9A–C). For Europe and North America (Fig. 9A), formations accumulated almost along a straight line and certainly they do not show any evidence of tailing off towards the present. For the rest of the world, new formations accumulated slowly from 1840 to 1940, but the rate of discovery has accelerated since then, as expected, and this effect is strongest when South Africa is excluded (cf. Fig. 9B–C). The discovery of new taxa tracks the taxon-bearing formations accumulation curve most

closely for the rest of the world as a whole (Fig. 9B), but lags a little for the rest of the world without South Africa (Fig. 9C) and lags a great deal for Europe and North America (Fig. 9A).

These differences are highlighted by the T:F ratios. For Europe and North America (Fig. 9D), the T:F ratio rose rapidly from 1820 to 1875 and has switched to a slower rate of increase since then. This might suggest that efforts since 1875 have included the discovery of new formations and new taxa in virgin territory, combined with re-exploration of already known formations from which new taxa are occasionally identified. For the rest of the world (Fig. 9E), the initial rapid rise in the T:F ratio, from 1840 to 1920, reaches a peak of 6.0–6.5 and then levels or falls slightly. This T:F ratio is twice that for Europe and North America, perhaps partly reflecting the fact that formations in those northern continents tend to be smaller than those in the rest of the world (maybe half the size, so yielding a T:F ratio of up to 2.9 for Europe and North America) or, perhaps less likely, that there is a great deal of difference in levels of synonymy and taxonomic revision between the two parts of the world. When South Africa is removed from the rest of the world (Fig. 9F), the T:F curve changes shape markedly, showing very little movement until the 1920s, when values began to rise steeply, with some steps, in the 1940s and 1980s, to a value of 3.5. This is more comparable with the Europe–North America ratio of 2.9, so may confirm that the Karoo is distorting the ‘rest of the world’ values (Fig. 9E) by including a small number of geographically vast formations that are each occupied by rather large faunas of early tetrapods, so raising the overall current T:F ratio to 6.09. Formations and species half-lives (Table 3) confirm these differences, being comparable between ‘old’ and ‘new’ lands, but being much more recent when the South African data are excluded from the ‘new’ lands. It would be hard to construct a purely sampling argument to explain the differences in T:F values between ‘old’ and ‘new’ continents (2.91 vs 6.09; Table 3; Fig. 9D–E) – the Karoo ‘formations’, really ‘assemblage zones’, each yielded 20–80 species (Nicolas and Rubidge 2010); this contributes to the largest of all T:F ratios (6.09) found in this study.

Dinosaurs show something similar. In Europe and North America (Fig. 10A), both formations and taxon discovery curves rise roughly in parallel, with an apparent asymptote beginning about 1930 and running to about 1980, when both curves accelerate. Interestingly, when these data were first explored (Benton 1998), it seemed that the curves of dinosaur discovery were truly reaching an asymptote, and it was suggested that this discovery curve pattern might indicate a means to estimate the ultimate global diversity of dinosaurs. No such luck. The acceleration in rates of discovery of dinosaurs in the ‘old’

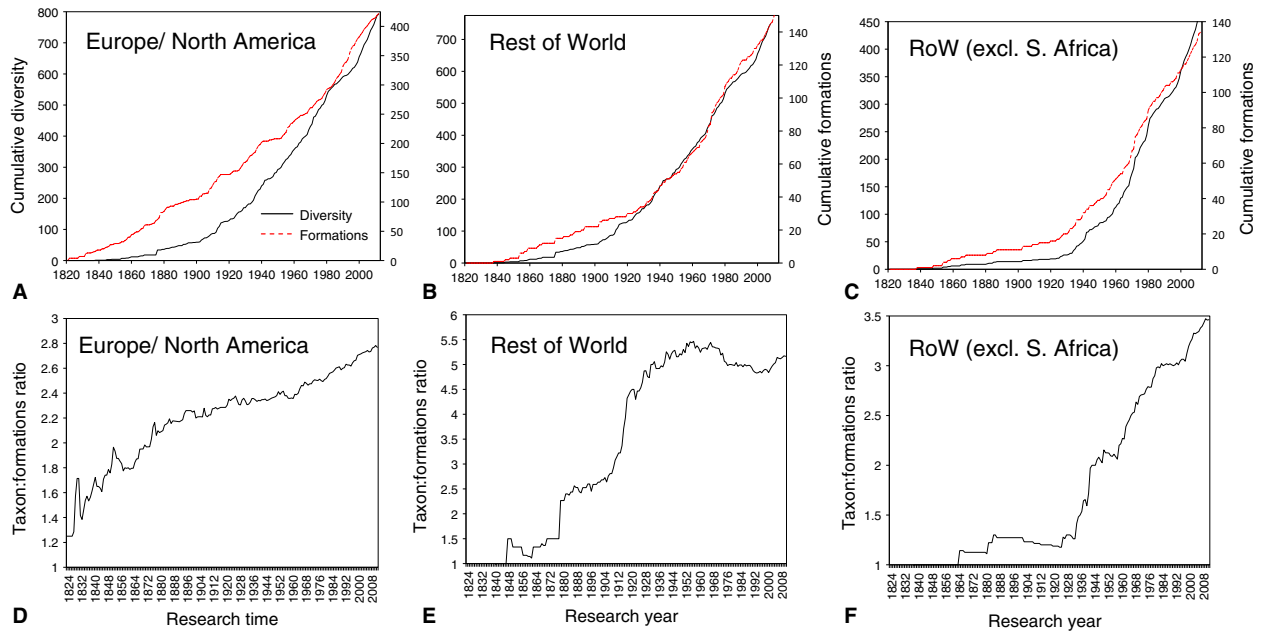
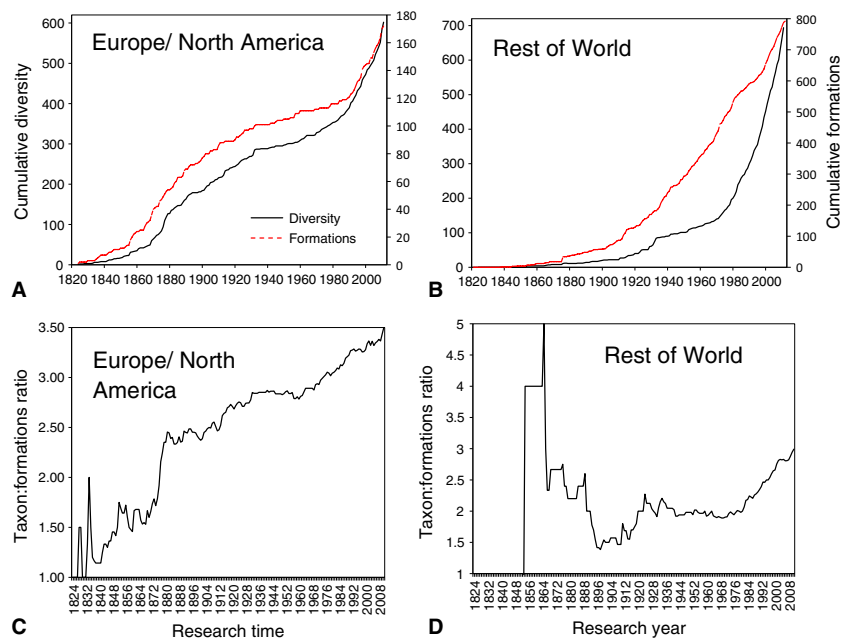


FIG. 9. Comparison of discovery curves and taxon:formation ratios for early tetrapods, as sampled in different geographical regions. Discovery curves (A–C) and T:F ratios (D–F), for Europe and North America (A, D), the rest of the world (B, E) and for the rest of the world excluding South Africa (C, F). These all show subtly different patterns of change through research time, perhaps reflecting aspects of differing sampling regimes more than original faunal compositions. Colour online.

FIG. 10. Comparison of discovery curves and taxon:formation ratios for dinosaurs, as sampled in different geographical regions. Discovery curves (A–B) and taxon:formation ratios (C–D), for Europe and North America (A, C) and the rest of the world (B, D). These show very different patterns of change through research time, showing a convergence on similar species-per-formation ratios, but following very different historic patterns, presumably documenting aspects of sampling intensity. Colour online.



lands is mimicked and exaggerated when the plot for the rest of the world is considered (Fig. 10B). Here, discoveries began in the 1850s, remained low until 1900 or 1920 and have then accelerated ever since, with cumulative new dinosaur-bearing formations keeping ahead of cumulative new dinosaurian taxa. Whereas the rate of discovery of new dinosaur-bearing formations has remained

steep but steady, the rate of naming new dinosaurs from the rest of the world shows an ever-accelerating, exponential curve (Fig. 10B). Some of this, surely, reflects the enthusiasm of the palaeontologists as much as reality, and doubtless many of the 50 or so new dinosaurs named each year since 2010 may turn out to be synonyms, *nomina dubia*, or *nomina nuda* (cf. Benton 2008a).

The T:F ratios through research time document these relationships. That for Europe and North America (Fig. 10C) is reminiscent of the T:F plot for early tetrapods from the same 'old' lands (Fig. 9D), showing a rapid rise from 1820 to 1880, and then a switch to a steady but slowly rising curve, indicating the slow acquisition of new dinosaur-bearing formations and a steady addition of small numbers of new taxa from already identified dinosaurian formations. For the rest of the world (Fig. 10D), the initial fluctuating values reflect small sample sizes, then show a plunge to rather low values from 1870 to 1900 and a fitful rise to a plateau from 1920 to 1980, followed by a rapid rise to the present day, quite unlike the equivalent early tetrapod plots (Fig. 9E–F). For dinosaurs, it seems that from 1870 to 1980 dinosaurian palaeontologists were behaving in a patchy fashion, at times adding new dinosaur-yielding formations, but those with rather rare dinosaurs (hence, overall falling T:F ratio), at others finding new formations and new dinosaurs at a one-to-one rate (hence, unchanging T:F ratio) and then, since 1980, finding new formations and new dinosaurs at a massively rising rate, but new taxa faster than new formations (hence, rising T:F ratio).

The half-life data (Table 3) bear this out, with an extraordinary difference between 'old' lands (where half the formations had been identified by 1914) and 'new' lands where this was achieved by 1986. These two figures are, respectively, the oldest and youngest formation half-lives discovered in the present study. The next oldest formations half-life (Table 3) is 1942, for Triassic–Jurassic dinosaurs, dominated, as noted earlier, by collections from Europe and North America, where few new formations are being identified. Likewise, the next youngest formations half-life was 1984, for Cretaceous dinosaurs as a whole, heavily dominated by discoveries outside Europe and North America.

DISCUSSION

Discovery curves and sampling

Discovery curves can indicate different sampling regimes in terms of the overall curve shapes and changing T:F ratios. For example, differing sampling is suggested by the fact that the Permian T:F ratio stabilized in 1940, but the Triassic T:F ratio continues to rise (Fig. 6). Permian formations and taxa continue to accumulate, but follow a straight line, whereas the Triassic curves are more concave, indicating rising discovery rates. The rising T:F ratio for the Triassic suggests that intensive search in existing tetrapod-bearing formations is yielding new taxa, so increasing the species-per-formation count.

A similar change in sampling is indicated for dinosaurs (Fig. 5F), where the sharp rise from 1970 onwards presumably reflects increased sampling efforts in the past 50 years, driving the count of genera recorded per formation from 2.5 to 3. The equivalent drop in synapsid ratios, from 5 to 4.5, over the same time period (Fig. 5C) might indicate the addition of many new formations with dinosaurs, but each yielding rather few new species per formation; this could be interpreted as evidence for an intensification of sampling in which palaeontologists are exploiting hitherto 'marginal' geological formations in which synapsid fossils are sparse, perhaps including the Late Triassic and Early Jurassic units that yield micro-remains of mammals.

The comparison of discovery curves and T:F ratios for Triassic–Jurassic and Cretaceous dinosaurs (Fig. 7) probably indicates different sampling regimes, reflecting the different states of search and knowledge. The data suggest that new Triassic–Jurassic formations found to contain dinosaurs are added only slowly, whereas much of the new discovery of taxa comes from searches within known dinosaur-bearing formations, so pushing the T:F ratio up rapidly. The constancy of T:F ratio for the Cretaceous (Fig. 7D) suggests closely coupled discovery of new dinosaurian genera and new dinosaur-bearing formations, as both totals increase rapidly and world coverage is improved.

The comparisons of 'well-sampled' regions (Europe and North America) with 'poorly sampled' regions (rest of the world) bear out some of these intuitions (Figs 9, 10). Apart from some sporadic signals in the data, perhaps reflecting small sample sizes in the early years, the rest-of-world data show that serious collecting began up to 100 years later in some cases (1920, rather than 1820) and that generally the rate of recovery of both new fossiliferous formations and new taxa are rising rapidly in the 'new' lands, but rising more slowly in the 'old' lands. In both cases, comparing the overall rates of increase between 'old' (Figs 9D, 10C) and 'new' (Figs 9E–F, 10D) lands confirms the null expectation, that the former are probably better sampled than the later. This is borne out also by the remarkable differences in half-lives, which are much younger for the 'new' lands than the 'old' lands (Table 3).

In all these cases, a simple data reduction is being considered. So long as the T:F ratio remains at or close to 1.0, then the direction of drive between formations \leftrightarrow taxa cannot be determined. Field palaeontologists identify one new fossiliferous formation for each new dinosaur they name when entering a new territory. Are they led to the new area by the hope of a new discovery, perhaps based on chance reports of 'giant bones' ('bonanza effect' of Raup (1977)); fossils \rightarrow formations), or do they search through a sedimentary basin and not find much until they chance upon a previously

unexplored formation, where their chances are improved by sampling over a wide area, heterogeneous rock successions that are divided into many formations ('bias effect' of Sheehan (1977); formations \rightarrow fossils)? Generally, the T:F ratio, for sporadically occurring tetrapods at least, will be 1.0 or higher.

When $T:F > 1.0$, then some formations at least have produced more than one taxon. This is the time when two collecting regimes can be identified: (1) the *exploratory regime*, when palaeontologists forever enter new territory and sample new formations until they find their first new taxon (and $T:F = 1.0$); and (2) the *re-sampling regime*, when they revisit known fossiliferous formations and find additional distinct taxa. In all cases considered here, formation count continues to rise, even in the well-sampled continents of Europe and North America, so the exploratory regime persists even in those parts of the world. But the rise in numbers of new fossiliferous formations identified in these 'old lands' has slowed in comparison with the 'new lands' where the addition of new formations accelerates.

At any time, differences in the T:F ratio can then reflect differences in the balance of exploratory vs re-sampling collecting modes, as well as differences in the original numbers of species per formation. The comparisons of the Permian and Triassic for early tetrapods and the Triassic–Jurassic and Cretaceous for dinosaurs (Figs 6, 7) provide striking evidence that T:F ratios can document radically different histories of knowledge. In the plot for Cretaceous dinosaurs (Fig. 7D), the constancy of the T:F ratio since 1850 is striking. The scope of knowledge about Cretaceous dinosaurs has increased massively since 1850, with thousands of new specimens and hundreds of new species collected from rich new deposits within hundreds of new fossiliferous formations that the Victorians could only have dreamt about. However, the T:F ratio has remained steady, indicating that dinosaur-bearing formations and genera of Cretaceous dinosaurs have been accumulating in concert virtually since the earliest days of dinosaur collecting. The same phenomenon is seen in the constancy of the T:F ratio for various subclades among early tetrapods (Fig. 5; Table 2), in some cases since 1870 (e.g. amphibians, dinosaurs). In such cases of constancy of the T:F ratio, the implication is that palaeontologists are continually seeking new ground, documenting new formations and new taxa at steady rates. At some point, palaeontologists may reach a point of reducing returns: the new formations they discover yield relatively few new taxa, so leading to a declining T:F ratio (e.g. Fig. 5C). In concert with this, palaeontologists may then return to actively seek fossils from the well-documented formations, both in the field and in museums, which can then drive the T:F ratio up (e.g. Figs 6D, 7A), although there is an upper limit to the T:F ratio, reflecting the mean numbers

of taxa that existed together within the region and time represented by a typical geological formation.

Formation counts as sampling metrics, or not

It is not clear then that the strict fossiliferous formation count (Benton *et al.* 2011) can be interpreted simply as a metric of sampling, as argued hitherto (Peters and Foote 2001; Smith 2001; Peters and Foote 2002; Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011; Lloyd 2012; Benson and Upchurch 2013). By following the palaeontologist into the field and observing how regional and global counts of new fossiliferous formations and new taxa accumulate in tandem, there is clearly some redundancy inherent in the data. This was argued by Benton *et al.* (2011), and several recent studies have now acknowledged this problem, applying a wider sampling proxy as well as, or instead of, a strict one (e.g. Upchurch *et al.* 2011; Pearson *et al.* 2013; Newham *et al.* 2014). In further detail, the coupling of the two signals at any time in research history is a form of operational redundancy, where formations and taxa may drive each other, and it is hard to disentangle whether new finds result from the bonanza or bias model. This uncertainty about drive direction might seem an arcane quibble, but it becomes important if the fossiliferous formations count is treated as a means of identifying times of variable sampling. Not identifying the drive direction, or whether the drive is bidirectional (redundancy) could lead to false assumptions. So, for example, dinosaurs show high palaeodiversity in the Campanian and Maastrichtian. Modelling studies (e.g. Barrett *et al.* 2009; Upchurch *et al.* 2011; Benson and Mannion 2012) have found that most of this dramatic increase can be accounted for by sampling, as represented by various formation count measures (strict and wider), and so a rapid palaeodiversity rise is then modelled as flat, or even as a decline in reality. If, however, the diversity really *was* high (and specimens are found in many fossiliferous formations, and specimen counts per formation were increasing), then it would be wrong to cut the diversity figures. The second interpretation (redundancy; diversity \rightarrow formations) is analogous to the species–area effect, where large islands contain many species in many locations, and small islands contain few species in fewer locations. 'Correcting' the diversity figures for large islands using locality count as a sampling metric would be nonsensical (Benton *et al.* 2011).

So long as the T:F ratio is equal to one, or remains constant at any value over long research time spans (e.g. Fig. 7D), the balance between new formations and new taxa is constant and both signals increase in tandem. Such cases reflect the exploratory phase of knowledge accumulation, with expansion of sampling as a proportion of the global total. In cases where the value of the T:F ratio rises

(e.g. Figs 6D, 7C, 9D–F, 10C, D), this typically reflects re-sampling of known formations, and so indicates improved sampling of known formations or regions.

Persistent differences in T:F ratios between different clades or different time bins could indicate substantial differences in sampling, whether by geological occurrence or by human effort, or they could indicate actual empirical differences in the mean diversities. The sampling explanations are most plausible in terms of differential fossilization, that, for example, one clade is undersampled with respect to another because of differences in habitat, body size, or skeletal robustness, or that tetrapods of a particular age are less well preserved because of an absence of suitable rock facies or because the animals were uniformly tiny in the Triassic, say, and much larger and heavier-boned in the Jurassic. Such explanations require independent evidence. Sampling differences in terms of human effort or taxonomic approach (e.g. splitting vs lumping) are probably generally less plausible; to claim, for example, that researchers of amphibians are more or less enthusiastic about collecting specimens or subdividing species than reptile researchers. Such differences in field and laboratory practice clearly exist – for example, nannofossil taxonomists working on different time intervals had different taxonomic concepts that resulted in different species to genus ratios (Lloyd *et al.* 2012) – but it would be hard to sustain an argument that such differences are general to the majority of researchers and over long spans of research time.

It is more likely that the bulk of such persistent differences in T:F ratios, by clade and by time bin, actually represent real, empirical differences. For example, the constant difference between T:F ratios for Permo-Triassic amphibians (2.45) and amniotes (4.03) would suggest relative faunal rarity of the former in comparison with the latter. Evidence for this is that body size distributions are not substantially different between Permo-Triassic amphibians and reptiles (both groups include a range of tiny to large forms), both clades are found in the same formations and horizons, and so mostly lived together, and all clades have been subject to substantial taxonomic revisions in recent years. Some Early Triassic faunas, for example those from Russia, are dominated by amphibians, confirming that it is possible to identify times when amphibians were more diverse and abundant than amniotes. These assumptions are borne out by independent, regional-scale collecting evidence. For example, in the Karoo basin of South Africa, temnospondyls (amphibians) comprise 0–20% of faunal samples, whereas synapsids comprise up to 100%; counts of genera range from 1 to 5 for amphibians and >35 for synapsids, based on hundreds to thousands of specimens in each of the assemblage zones (Nicolas and Rubidge 2010; Irmis and Whiteside 2012;

Irmis *et al.* 2013; Smith and Botha-Brink 2014). Likewise, in the extensive Permo-Triassic redbed successions of Russia, amphibians are generally less abundant in individual svitas than amniotes, except in the earliest Triassic, also based on the sampling over 80 years of thousands of specimens from hundreds of localities (Tverdokhlebov *et al.* 2003, 2005; Benton *et al.* 2004).

Further, in the comparisons of ‘well-sampled’ continents (Europe, North America) and less-well-sampled continents (rest of the world), the effects of sampling on the T:F ratio were evident, and the findings confirmed null expectations (Figs 8–10). However, again, there may be an interplay of empirical and sampling signals. The Cretaceous dinosaur record is still growing fast, largely as a result of continuing high rates of exploration in new regions, whereas the Triassic–Jurassic dinosaur record shows a long-term steadily rising T:F ratio, indicating a mix of exploration and re-sampling regimes (Fig. 7). In other words, the contrast between the two divisions of dinosaurian history, which was noted before (Benson and Mannion 2012), is actually more complex, reflecting differences in sampling regime and in the empirical biological pattern. Sampling regimes differ in each time division, with re-sampling of known formations dominating the growth of knowledge for the Triassic–Jurassic, and exploration of new territory dominating for the Cretaceous. But it is not all sampling: a key empirical difference is that there are many more dinosaur-bearing formations and dinosaurs in the Cretaceous than in the Triassic–Jurassic, and those numbers are growing faster in the Cretaceous than in the Triassic–Jurassic. The persistently higher T:F ratio of 3.6 for the Cretaceous (cf. Triassic–Jurassic, 2.7) shows that, despite intense efforts in re-sampling known formations, and seeking new ones, throughout the well-trodden outcrops of Europe and North America, there is a substantial and real difference that is not sampling: dinosaurs were simply much more abundant and diverse in the Cretaceous than in the Triassic–Jurassic.

These new observations can now enrich earlier studies in which dinosaurian palaeodiversity was modelled against varying formation counts (e.g. Barrett *et al.* 2009; Mannion *et al.* 2011; Upchurch *et al.* 2011; Benson and Mannion 2012). Barrett *et al.* (2009, fig. 2), for example, showed remarkably small residuals for Cretaceous dinosaurs, a corollary of their finding of a high correlation between counts of dinosaurian genera and dinosaur-bearing formations. They noted (Barrett *et al.* 2009, p. 2667) that ‘Strong statistically robust correlations demonstrate that almost all aspects of ornithischian and theropod diversity curves can be explained by geological megabiases’ (Kowalewski and Flessa 1996). This analysis, and others, did not take account of the underlying differences in T:F ratios that we show here, and so substantially underestimated Cretaceous dinosaurian diversity, which, as noted here, was about

twice as high as that in the Triassic–Jurassic. It turns out that all the cumulative indicators of sampling, such as counts of collections, localities and dinosaur-bearing (even tetrapod-bearing) formations, increase in concert, and their close correlations indicate operational redundancy of signals (Benton *et al.* 2011; Dunhill *et al.* 2014a).

The discussion above has focused on strict fossiliferous formation counts (e.g. dinosaur-bearing formations vs dinosaur diversity). Wider fossiliferous formation counts have been used in a number of studies (e.g. Benton *et al.* 2011, 2013a; Upchurch *et al.* 2011; Pearson *et al.* 2013; Newham *et al.* 2014), and these have the advantage of allowing for non-sampling of the taxon under study. Here, I have not explored the relationships between strict and wider fossiliferous formation counts, nor between fossiliferous and unfossiliferous formation counts. By definition, the entire scope of the rock record is underestimated if total formation counts (= fossiliferous + unfossiliferous formations) are not used. As an example, it would be worth knowing whether the regional or global measure of exposure/rock volume/barren formations count changed by an order of magnitude between any pair of time bins in a palaeodiversity analysis. In their search for fossils, palaeontologists have access to fossiliferous and barren rock formations, and total formations count is presumably closest to reflecting the independent metrics of the rock record, such as map area, rock volume or exposure counts. However, for reasons noted above and in Benton *et al.* (2011), formations are human constructs, and so they probably introduce a substantial skewing factor into attempts to estimate the rock record. This has yet to be demonstrated by numerical analysis.

Formation counts as a basis for modelling bias

These observations chime broadly with the use of strict formation counts in modelling bias on the simple assumption that the two signals are linked (Peters and Foote 2001, 2002; Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011; Benson and Mannion 2012; Benson and Upchurch 2013; Newham *et al.* 2014). The aim is to use the relationship between formations and taxa, assuming that formations count wraps up many geological and human aspects of sampling, to identify times when the palaeodiversity signal cannot be explained by sampling, and so might reflect a true, biological signal. However, as Smith and McGowan (2007) and Lloyd (2012) noted in presenting the bias-modelling method, there is a concern that the detrending process that is used to remove bias may remove some true biological signal at the same time.

The problem may be worse than suggested by Lloyd (2012) and others: when strict formation count is used as

the indicator of bias, then probably both the main trend ('bias model') and the detail (residuals) both encompass complex sampling and empirical signals. This violates the assumption behind the bias-modelling approach, that bias and empirical signal can be separated. Variations in the T:F ratio have been shown here to reflect original faunal proportions (important if members of a clade showed varying local and regional diversities through time) and suggested also to reflect changing overall clade size, regional and global changes in species distributions relating to climate change, and differing sampling regimes. Therefore, an overall rising trend through geological time probably combines a measure of improved sampling and truly higher diversity through time (compare the collecting records of Triassic–Jurassic vs Cretaceous dinosaurs in which Cretaceous dinosaurian diversity is empirically higher than Triassic–Jurassic dinosaurian diversity; Fig. 7). It is not clear then that any of the cumulative collecting proxies, such as fossiliferous formation counts, can provide a clear distinction in modelling between sampling and true signal.

The corollary is probably also true. Whereas it is argued that the residuals should then reflect predominantly biological signals (Lloyd 2012), this need not be the case. The residuals are not necessarily free of sampling, in that divergences from the steady T:F ratio, or overall trend, can include sampling effects, or reflect sampling effects entirely. For example, a positive residual peak in a particular geological time bin might result from good sampling in the form of an unusual site of exceptional preservation (one formation, many taxa), a problem noted by Lloyd (2012), and a negative residual in another geological time bin might reflect a shortage of rocks of that age, and so poor sampling. To give a concrete example, the mismatch of diversity peaks (Barremian–Aptian) and formation peaks (Aptian–Cenomanian) in the Early Cretaceous (Fig. 2B) can be explained by a genuine rise in diversity, but this comparison masks the occurrence of the Yixian Formation of the Jehol Group, an enormously geographically extensive geological formation that has famously produced thousands of dinosaur and bird specimens. Strong sampling effects produced by Lagerstätten such as the Yixian Formation are missed by all variants of formation counts, but this issue has been recognized and methods to ameliorate or remove the effect have been developed (e.g. Butler *et al.* 2009).

Modelling sampling bias on the basis of fossiliferous formation counts may be inappropriate because the modelled trend and the residuals are not what they are claimed to be. The modelled trend is not necessarily predominantly bias, but might be dominated by real biological signal (e.g. a diversifying clade), and the residuals need not reflect surplus empirical signal not accounted for by the modelled trend – they could just as well be

entirely caused by quirks of sampling or by real extinctions and diversifications. The extraordinary outcome of many bias-modelling analyses can be seen with Dinosauria – the substantially rising diversity in the Cretaceous is damped, and the apparent dramatic rise at the end of the Cretaceous is turned into a decline (Barrett *et al.* 2009; Lloyd 2012; Upchurch *et al.* 2011). The implication of accepting this modification to the diversity signal is that the empirical record of expansion of Dinosauria is incorrect. The expansion of Dinosauria through the Mesozoic, and especially in the Cretaceous occurred in several ways: wider geographical occurrence (more formations and localities over more regions and continents); more species per fauna (rising from 2–5 dinosaurs per fauna/formation in the Triassic to 20–50 in the Late Cretaceous); and more Baupläne, represented by major lineages (22 in Late Triassic; 130 in the Late Cretaceous; Benson *et al.* 2014, fig. 3). These aspects could be missed by a modelling approach that assumes that a fossiliferous formation count provides a reliable model of sampling, and underlying evidence for global expansion can be lost.

The point is not that sampling can be ignored, but that the current modelling approaches that use formation count to guide the estimation of bias, do not achieve what is claimed. These are the bugbears of all attempts to remove sampling from palaeodiversity time series – it is hard to find an objective way to determine whether diversity dips, or negative residuals, reflect empirical rarity of taxa or poor sampling (Wignall and Benton 1999), or whether diversity peaks or positive residuals reflect empirical abundance or good sampling. Far better to use independent metrics, such as map areas or rock volume metrics, as recommended by Smith and McGowan (2007). Further, measures of effort and measures of rock heterogeneity could be incorporated to provide an *a priori* more reasonable sampling-dominated signal against which to model the empirical fossil record.

Bias, common cause or redundancy?

There is probably no single, decisive test among the bias, common cause and redundancy hypotheses. This reflects the complexity of relationships among all the metrics, and the fact that all are in play to a greater or lesser extent in any particular case study. However, it is important to resolve which dominates to determine whether, for example, a pattern of raw fossil diversity, or palaeodiversity, means anything biological or not.

There has been a genuine divergence of opinion over the use of sampling metrics, where some (e.g. Crampton *et al.* 2003; Smith 2007; Benton *et al.* 2011) have argued that sampling metrics should be demonstrably independent of the signal they seek to regulate, and

others (e.g. Peters and Foote 2001; Barrett *et al.* 2009; Benson and Mannion 2012) have used measures such as counts of fossiliferous formations, localities or collections as that yardstick. In ecological sampling (Lohr 2009; Albert *et al.* 2010; Jensen and Bourgeron 2012), the object of sampling is usually a fixed entity, such as the flora of an island or an elephant herd through time, and the means of sampling is according to fixed increments (e.g. a 1 m² quadrat; 1 day or 10 person-weeks of effort) that are independent of the population being sampled. The palaeontological sampling metrics of numbers of collections, localities or formations fail the latter two criteria in that they are not fixed or comparable sampling aliquots (each collection, locality or formation may be of very different magnitude; formations, for example, vary over eight orders of magnitude; Benton *et al.* 2011) and their variations through time are not independent of the population being sampled, namely the recorded palaeodiversity curve. It is important to note that the objections to palaeontological sampling metrics on the grounds of their great variability in dimensions would not apply if they were scattered evenly through the time bins, such that the means and standard deviations were constant through the time range of interest; such comparative analyses have yet to be done. Should they not be evenly scattered through the time bins of interest, then the approach would be rather like observing an elephant herd by using observation spans of different timing (say, 1 h on day 1, 12 h on day 2, 3 min on day 3) and varying the observation time dependent on the number of elephants first spotted each day. A first, detailed, regional study (Dunhill *et al.* 2014a) has confirmed these concerns and shown that collections and formations show bidirectional information transfer (= redundancy) with palaeodiversity, and only outcrop area can drive palaeodiversity.

There are several possible, empirical, ways to determine the plausibility of the sampling or redundancy model in any palaeontological example, whether global or regional, and six of these are outlined below.

Test of directionality. Most statistical techniques, such as correlation/time series model fitting, indicate whether two variables covary, to what extent, and whether the relationship is positive or negative, but they do not identify the *directionality* of the relationship, that is, which variable *drives* the other, or whether the relationship is two-way. Information transfer techniques (Hannisdal 2011) can determine directionality, if it exists. In an initial study of the global marine fossil record, Hannisdal and Peters (2011) found that much of the observed covariation between patterns of sedimentation and palaeodiversity depended on mutual responses to interacting Earth systems, not on sampling biases. Further, in the first study of a regional case using such methods, Dunhill *et al.*

(2014a) showed that in the British fossil record, marine outcrop area contains a signal useful for predicting changes in diversity, collections and formations, and terrestrial outcrop area contains a signal useful for predicting formations. On the other hand, collection and formation counts were information redundant with fossil richness, characterized by symmetric, bidirectional information flow. This demonstrates that collection and formation counts are redundant with palaeodiversity, and so cannot be used as evidence to correct, or even rank, time bins with poor or good sampling. These two studies are conclusive and reject the bias model as an explanation for covariation of formation counts and palaeodiversity, but additional regional and global case studies will be required to confirm these findings.

Correlation of formation counts with other sampling proxies. If formation count is a synoptic sampling metric, it ought to correlate with other sampling metrics, such as outcrop area, rock volume or number of sections. In the North American rock record, COSUNA sedimentary rock sections correlate with formation counts (Peters 2005), a finding that has been widely cited as evidence to use the formation count as a single, all-encompassing sampling metric (Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011; Upchurch *et al.* 2011; Benson and Upchurch 2013). In addition, Lloyd and Friedman (2013) found considerable similarities between three sampling proxies for British fossil fishes, fossiliferous formations, localities and, importantly, map grid occupancy (a measure of outcrop area). However, in their study of the dinosaurian fossil record, Upchurch *et al.* (2011) also found correlations between formation counts and some rock record proxies for North America and Europe, but there was no correlation between regional measures of outcrop area for North America or Europe with a global dinosaur-bearing formation count. Likewise, Benson *et al.* (2013) found similar results for the Cretaceous tetrapod fossil record, with generally good correlation of formation counts and rock volume metrics in North America, but less so in Europe (other continents were not assessed because of the absence of reliable rock volume metrics). In a study of the first half of tetrapod evolution (Benton *et al.* 2013a), various versions of the formation count proxy did not correlate with rock record metrics, at global scale, and the only reliable correlations were between families of co-dependent metrics from the same databases (including correlation of various Paleobiology Database (Alroy *et al.* 2001) measures with each other, and correlation of various Macrostrat database (Peters and Heim 2010) measures with each other). Further, Crampton *et al.* (2003) found no correlation between outcrop area and formation counts in their New Zealand case study, and Dunhill *et al.* (2014b) found mixed results in a study of the Triassic–Jurassic of

the UK. Further, each supposed sampling proxy is in itself highly variable (compare different widely used metrics for sea level, rock volume, map areas) and the case could be argued either way. In their detailed regional study of the British rock and fossil records, Dunhill *et al.* (2014a) found that outcrop area contains a signal useful for predicting formation counts, and so suggested that outcrop area, if appropriately measured, would be an appropriate sampling proxy, but formation count probably would not.

Comparison of multiple studies. The ‘residuals method’ of Smith and McGowan (2007) generates a model in which a fossil record sampling proxy, such as formation count or global outcrop area, is used to estimate the amount of taxonomic diversity expected for that amount of sampling if diversity were equal in all time intervals. Positive and negative residuals indicate variations from the predicted curve that cannot be explained by the sampling proxy, and these mean either that palaeodiversity was truly much higher or lower than the norm (biological signal) or that sampling was much poorer or better than the norm (sampling). The residuals cannot be used simply as an indicator of under- or over-sampling, but different sampling metrics applied to the same fossil records might be expected to yield comparable results if it is valid to assume that the bulk of the covariation between palaeodiversity and formations time series is to be interpreted as evidence that the former is explained by the latter. In a broad-scale study, Smith *et al.* (2012) showed good congruence between residuals from a map-area modelled curve of marine animals and a shareholder quorum subsampled curve, and concluded that their mutual congruence confirmed the validity of both approaches and, importantly, identified times when diversity is unexplained by the surviving rock area model. On the other hand, some recent studies of Mesozoic tetrapods have yielded mixed results (Table 4). Dinosaurs have been analysed twice, using variants of the dinosaur-bearing formation count (Barrett *et al.* 2009; Upchurch *et al.* 2011), and they show rather different patterns; there is good agreement between the two studies of Ornithischia, but not for Theropoda and Sauropodomorpha. On the other hand, good concordance is seen between the residuals analyses of the fossil records of pterosaurs and of birds, where several episodes of significantly negative residuals coincide, suggesting that the sampling proxies (variants of formation counts in each case) explain the palaeodiversity signals similarly. The bird–pterosaur examples are unusual because most of their fossil records are dominated by Lagerstätten, and specimens of both are found in these same sites of exceptional preservation, so perhaps explaining the similar signals. These results overall suggest complex interplay between formation counts and palaeodiversity signals, but on their own cannot test

TABLE 4. Summary of times of significantly negative residuals from the formations–palaeodiversity relationship, from different analyses.

	Ornithischians		Theropods		Theropods		Sauropodomorphs		Sauropodomorphs		Pterosaurs	Birds	Mammals – wide		Mammals – narrow
	1	2	1	2	1	2	1	2	2	2					
Late Cretaceous											×	×			
Maastrichtian											×	×			
Campanian											×	×			
Santonian															
Coniacian													×		
Turonian													×		
Cenomanian					×					×					
Albian															
Early Cretaceous															
Aptian															
Barremian										×					×
Hauterivian						×				×					×
Valanginian						×				×					×
Berriasian						×				×					×
Tithonian										×					
Late Jurassic															
Kimmeridgian															
Oxfordian						×									
Mid. Jurassic															
Callovian		×													
Bathonian	×														
Bajocian	×														
Aalenian	×	×													
Toarcian	×	×							×						
Early Jurassic															
Pliensbachian		×													
Sinemurian			×												
Hettangian															
Late Triassic															
Rhaetian									×				×		×
Norian													×		×
Carnian															

Data are summarized from Barrett *et al.* (2009) for Ornithischia 1, Theropoda 1 and Sauropodomorpha 1; Upchurch *et al.* (2011) for Ornithischia 2, Theropoda 2 and Sauropodomorpha 2; Butler *et al.* (2009) for pterosaurs; Brocklehurst *et al.* (2012) for birds; and Newham *et al.* (2014) for mammals ('Mammals – wide' is compared with an all-terrapods formation count; 'Mammals – narrow' with a mammals-only formation count). × indicates a negative residual below the 95% confidence envelope.

between the bias and redundancy models. They do suggest the need for caution, however, if the palaeodiversity and sampling metric comparisons show some commonalities, but can also vary according to the particular data sets and sampling proxies employed.

Part and whole. There might be a relationship between ‘wide’ and ‘narrow’ formation counts, for example, between a tetrapod-bearing formation count and a dinosaur-bearing formation count. Covariation of both would imply shared probability of preservation (equivalent relative abundances, habitats, distribution of hard parts, size) and study (equivalent levels of interest in terms of person-hours and publications, equivalent rationale for distinguishing and naming new taxa). The use of wide formation counts has been recognized as important (e.g. Barrett and Upchurch 2005; Benton *et al.* 2011, 2013a; Benson *et al.* 2013; Newham *et al.* 2014) as a minimal means of avoiding a strict one-to-one formation-taxon redundancy, especially in cases where each species is typically represented by a single skeleton or species from a single locality or formation. For example, Fröbisch (2008) found a very exact correspondence of formation and locality counts with Permo-Triassic anomodont taxa, but when compared to an all-tetrapod metric (Benton *et al.* 2011; Fröbisch 2013), the tight linkage disappeared. Further, in a study of the Mesozoic mammal fossil record (Newham *et al.* 2014), two versions of the formation count (all-tetrapod, and mammal-only) showed agreement in highlighting significantly negative residuals in the Late Triassic and Early Cretaceous (the latter slightly offset), but only the former showed such a dip in the Late Cretaceous also. These ideas need further exploration, using different groups of taxa, and further revised formation count metrics. Importantly, the wide formation count metrics allow for the detection of non-preservation (Benton *et al.* 2013a): rare taxa might just not be sampled, and so their host formations would be excluded from a narrow formation count, but would be highlighted by a wide formation count that incorporated information about the occurrence of other clades. Benson *et al.* (2013) concluded, from their study of global and continental-scale aspects of the Cretaceous fossil record of tetrapods that ‘the absence of strong statistical relationships between tetrapod sampling proxies from different continental areas suggests that there is no unified ‘global’ sampling signal for terrestrial tetrapods.’ These findings confirm the complexity of the relationship between regional and global scales in understanding empirical palaeodiversity and sampling errors (McGowan and Smith 2008).

Improvement through time. If formation count is a metric of sampling and includes some, or many, human aspects, then sampling of any clade worldwide or regionally might

be expected to improve through time. As shown above, the term ‘improve’ might be seen in two contexts, reflecting the two collecting modes of exploration and re-sampling. We found that the rates of discovery of new formations and new dinosaurs are rising rapidly outside Europe and North America, indicating massively increasing knowledge, but there is still no impression of when these discoveries will slow down, as they have in Europe and North America. Therefore, sampling of dinosaurs has clearly improved massively over the years in that so many more taxa and source formations are known now than in the past, but whether one can say that the mean level of sampling of the currently known sum total of dinosaur-bearing formations has improved overall is a different matter. Our comparisons of fossil sampling between well-sampled continents (Europe, North America) and poorly sampled continents (rest of world) confirmed the dominance of re-sampling known formations in the former and exploration of new territory in the latter (Figs 8–10).

Second discoveries. As we have seen, formation counts and taxon counts rise in lockstep in cases where specimens and taxa are rare: this is a pure example of bidirectional formation–taxon redundancy (Benton *et al.* 2011, 2013a; Dunhill *et al.* 2014a). If those first discoveries from each formation are removed, this close formation–taxon redundancy would also be removed; this can be done by simply exploring the portions of the T:F plots (e.g. Figs 5, 6C–D, 7C–D, 9D–F, 10C–D) that lie above the value, T:F = 1.0. The T:F value is determined by the original relative abundances of different taxa, some being rare and others common within their faunas or formations so the slope of that relationship through time is most interesting. Unchanging T:F values through time, as shown most strikingly for Cretaceous dinosaurs (Fig. 7D) reflect steady levels of sampling. A rising T:F ratio, as seen in many cases (Figs 5D–F, 6D, 7C, 9F, 10C), indicates improved sampling (more taxa per formation through research time), presumably eventually tending to the T:F ratio that reflects original abundance at some point of saturation. A falling T:F ratio may indicate simply that highly fossiliferous (= large) formations were sampled first, and newly found formations are smaller or less fossiliferous; even so, a falling ratio does not indicate poorer sampling. All of these conclusions are based on narrative assumptions and cannot be explicitly tested with a single statistical analysis.

CONCLUSION

The strong correlations commonly found between taxon counts and counts of formations bearing those taxa, both in previous work (Peters and Foote 2001, 2002; Fröbisch

2008; Barrett *et al.* 2009; Butler *et al.* 2009; Benson *et al.* 2010; Mannion *et al.* 2011; Benson and Upchurch 2013), and here (Table 1), reflect a mix of several signals, including redundancy (each dinosaur-bearing formation was added to the roster at the same time as a new dinosaur genus), empirical signal (as dinosaur diversity rose and fell through geological time, the numbers of dinosaur-bearing formations rose and fell) and sampling signal (appropriate rocks are present to different extents through geological time). The close covariation of taxon and formation counts confirms something about how palaeontologists work. Discovery is the dream, and palaeontologists have for centuries trekked over suitable rocks looking for fossils. When a fossil is found and turns out to be a new species, both the taxon and the formation have been discovered, and they are added to the global rosters for that fossil group. When considering how palaeontologists operate in the field, as noted by Raup (1977), it is evident that fossil discovery mainly drives (fossiliferous) formation discovery, and this is borne out by the analyses of discovery patterns of early tetrapods and of dinosaurs in this paper.

The key question, however, is whether bias in the fossil record seriously distorts our understanding of the history of life. One series of studies seemed to suggest that the answer was no (Maxwell and Benton 1990; Sepkoski 1993; Benton 2008a; Bernard *et al.* 2010). In these, different segments of the fossil record were sampled through research time to determine how the growing data set affected understanding of macroevolutionary patterns. The finding was that, despite massive effort by palaeontologists and considerable expansion of their fields of sampling to cover previously untouched continents, many macroevolutionary patterns remain little changed through research time. The conclusion of these studies was that palaeontologists made equal efforts across time bins and across taxonomic groups, both to collect more material and to study the existing materials, and new discoveries rarely change the overall macroevolutionary pattern dramatically. However, these studies did not test whether the invariant palaeodiversity curves really represented the truth or not; continued collecting could simply enhance entrenched biases of preservation.

Comparisons of phylogeny and fossil record may be more fruitful in that phylogenies can be compiled largely independently of fossil occurrence data (Smith 2007). Such comparisons often show good congruence, suggesting that at least the order of appearance of fossils in the rocks is generally correct (Norell and Novacek 1992; Benton *et al.* 2000). When such comparisons are made through research time, they can show overall improvements in knowledge as predicted gaps in the fossil record (e.g. ghost ranges and Lazarus gaps) are filled (Benton and Storrs 1994). Results

from phylogenetic comparisons may be mixed (Tarver *et al.* 2011): substantial expansions of data through research time have not fundamentally rewritten understanding of diversity through time or phylogeny for primates, but phylogenetic knowledge of dinosaurs has improved with the addition of new taxa to the phylogenetic trees.

Acknowledgements. I thank Alex Dunhill, Neil Brocklehurst, Bjarte Hannisdal and Graeme Lloyd for very helpful comments on earlier drafts. I am especially grateful to Andrew Smith for his extensive work to help clarify the paper. This work was supported by NERC grant NE/C518973/1.

DATA ARCHIVING STATEMENT

Data for this study are available in the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.29j8d>

REFERENCES

- ALBERT, C. H., YOCCOZ, N. G., EDWARDS, T. C. and GRAHAM, C. H. 2010. Sampling in ecology and evolution – bridging the gap between theory and practice. *Ecography*, **33**, 1028–1037.
- ALLISON, P. A. and BRIGGS, D. E. G. 1993. Exceptional fossil record: distribution of soft-tissue preservation through the Phanerozoic. *Geology*, **21**, 527–530.
- ALROY, J. 2002. The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 3706–3711.
- 2010a. The shifting balance of diversity among major marine animal groups. *Science*, **329**, 1191–1194.
- 2010b. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaentology*, **53**, 1211–1235.
- MARSHALL, C. R., BAMBACH, R. K., BEZUSKO, K., FOOTE, M., FURSICH, F. T., HANSEN, T. A., HOLLAND, S. M., IVANY, L. C., JABLONSKI, D., JACOBS, D. K., JONES, D. C., KOSNIK, M. A., LIDGARD, S., LOW, S., MILLER, A. I., NOVACK-GOTTSHALL, P. M., OLSZEWSKI, T. D., PATZ-KOWSKY, M. E., RAUP, D. M., ROY, K., SEPKOSKI, J. J., SOMMERS, M. G., WAGNER, P. J. and WEBBER, A. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 6261–6266.
- BARRETT, P. M. and UPCHURCH, P. 2005. Sauropodomorph diversity through time: paleoecological and macro evolutionary implications. 125–156. In CURRY-ROGERS, K. A. and WILSON, J. A. (eds). *The sauropods: evolution and paleobiology*. University of California Press, Berkeley, 363 pp.
- MCGOWAN, A. J. and PAGE, V. 2009. Dinosaur diversity and the rock record. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 2667–2674.

- BEBBER, D. P., MARRIOTT, F. H. C., GASTON, K. J., HARRIS, S. A. and SCOTLAND, R. W. 2007. Predicting unknown species numbers using discovery curves. *Proceedings of the Royal Society of London B: Biological Sciences*, **274**, 1651–1658.
- BENSON, R. B. J. and MANNION, P. D. 2012. Multi-variate models are essential for understanding vertebrate diversification in deep time. *Biology Letters*, rsbl20110460. doi:10.1098/rsbl.2011.0460
- and UPCHURCH, P. 2013. Diversity trends in the establishment of terrestrial vertebrate ecosystems: Interactions between spatial and temporal sampling biases. *Geology*, **41**, 43–46.
- BUTLER, R. J., LINDGREN, J. and SMITH, A. S. 2010. Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 829–834.
- MANNION, P. D., BUTLER, R. J., UPCHURCH, P., GOSWAMI, A. and EVANS, S. E. 2013. Cretaceous tetrapod fossil record sampling and faunal turnover: implications for biogeography and the rise of modern clades. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 88–107.
- CAMPIONE, N. E., CARRANO, M. T., MANNION, P. D., SULLIVAN, C., UPCHURCH, P. and EVANS, D. C. 2014. Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. *PLoS Biology*, **12** (5), e1001853.
- BENTON, M. J. 1995. Diversification and extinction in the history of life. *Science*, **268**, 52–58.
- 1998. The quality of the fossil record of vertebrates. 269–303. In DONOVAN, S. K. and PAUL, C. R. C. (eds). *The adequacy of the fossil record*. Wiley, New York, 322 pp.
- 2008a. How to find a dinosaur, and the role of synonymy in biodiversity studies. *Paleobiology*, **34**, 516–533.
- 2008b. Fossil quality and naming dinosaurs. *Biology Letters*, **4**, 729–732.
- 2012. No gap in the Middle Permian record of terrestrial vertebrates. *Geology*, **40**, 339–342.
- 2015. Data from: Palaeodiversity and formation counts: redundancy or bias? *Dryad Digital Repository*. doi:10.5061/dryad.29j8d
- and STORRS, G. W. 1994. Testing the quality of the fossil record: paleontological knowledge is improving. *Geology*, **22**, 111–114.
- WILLS M. A. and HITCHIN, R. 2000. Quality of the fossil record through time. *Nature*, **403**, 534–537.
- TVERDOKHLEBOV, V. P. and SURKOV, M. V. 2004. Ecosystem remodelling among vertebrates at the Permian-Triassic boundary in Russia. *Nature*, **432**, 97–100.
- DUNHILL, A. M., LLOYD, G. T. and MARX, F. G. 2011. Assessing the quality of the fossil record: insights from vertebrates. *Geological Society, London, Special Publications*, **358**, 63–94.
- RUTA, M., DUNHILL, A. M. and SAKAMOTO, M. 2013a. The first half of tetrapod evolution, sampling proxies, and fossil record quality. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 18–41.
- — — — — 2013b. Data from: the first half of tetrapod evolution sampling proxies, and fossil record quality. *Dryad Digital Repository* doi:10.5061/dryad.44b50
- BERNARD, E. L., RUTA, M., TARVER, J. E. and BENTON, M. J. 2010. The fossil record of early tetrapods: worker effort and the end-Permian mass extinction. *Acta Palaeontologica Polonica*, **55**, 229–239.
- BROCKLEHURST, N. and FRÖBISCH, J. 2014. Current and historical perspectives on the completeness of the fossil record of pelycosaurian-grade synapsids. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **399**, 114–126.
- UPCHURCH, P., MANNION, P. D. and O’CONNOR, J. 2012. The completeness of the fossil record of Mesozoic birds: implications for early avian evolution. *PLoS One*, **7** (6), e39056. doi:10.1371/journal.pone.0039056
- BUTLER, R. J., BARRETT, P. M., NOWBATH, S. and UPCHURCH, P. 2009. Estimating the effects of sampling biases on pterosaur diversity patterns: implications for hypotheses of bird/pterosaur competitive replacement. *Paleobiology*, **35**, 432–446.
- CAIN, S. A. 1938. The species–area curve. *American Midland Naturalist*, **19**, 573–581.
- COSTELLO, M. J., WILSON, S. and HOULDING, B. 2012. Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Systematic Biology*, **61**, 871–883.
- CRAMPTON, J. S., BEU, A. G., COOPER, R. A., JONES, C. M., MARSHALL, B. and MAXWELL, P. A. 2003. Estimating the rock volume bias in paleobiodiversity studies. *Science*, **301**, 358–360.
- DUNHILL, A. M., BENTON, M. J., NEWELL, A. J. and TWITCHETT, R. J. 2013. Completeness of the fossil record and the validity of sampling proxies: a case study from the Triassic of England and Wales. *Journal of the Geological Society*, **170**, 291–300.
- HANNISDAL, B. and BENTON, M. J. 2014a. Disentangling rock record bias, common cause and redundancy in the fossil record. *Nature Communications*, **5**, 4818. doi:10.1038/ncomms5818
- BENTON, M. J., TWITCHETT, R. J. and NEWELL, A. J. 2014b. Testing the fossil record: sampling proxies and scaling in the British Triassic-Jurassic. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **404**, 1–11.
- FASTOVSKY, D. E., HUANG, Y., HSU, J., MARTIN-McNAUGHTON, J., SHEEHAN, P. M. and WEISHAMPEL, D. B. 2004. Shape of Mesozoic dinosaur richness. *Geology*, **32**, 877–880.
- FRÖBISCH, J. 2008. Global taxonomic diversity of anomodonts (Tetrapoda, Therapsida) and the terrestrial rock record across the Permian-Triassic boundary. *PLoS One*, **3**, e3733.
- 2013. Vertebrate diversity across the end-Permian mass extinction – separating biological and geological signals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 50–61.
- GRADSTEIN, F. M., OGG, G. and SCHMITZ, M. 2012. *The geologic time scale 2012*. Elsevier, Amsterdam, 1176 pp.

- HANNISDAL, B. 2011. Non-parametric inference of causal interactions from geological records. *American Journal of Science*, **311**, 315–334.
- and PETERS, S. E. 2011. Phanerozoic Earth system evolution and marine biodiversity. *Science*, **334**, 1121–1124.
- IRMIS, R. B. and WHITESIDE, J. H. 2012. Delayed recovery of non-marine tetrapods after the end-Permian mass extinction tracks global carbon cycle. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 1310–1318.
- and KAMMERER, C. F. 2013. Non-biotic controls of observed diversity in the paleontologic record: an example from the Permo-Triassic Karoo Basin of South Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 62–77.
- JENSEN, M. E. and BOURGERON, P. S. 2012. *A guidebook for integrated ecological assessments*. Springer, New York, 549 pp.
- KOWALEWSKI, M. and FLESSA, K. W. 1996. Improving with age: the fossil record of lingulide brachiopods and the nature of taphonomic megabiases. *Geology*, **24**, 977–980.
- LLOYD, G. T. 2012. A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves: new support for declining Cretaceous dinosaur richness. *Biology Letters*, **8**, 123–126.
- and FRIEDMAN, M. 2013. A survey of palaeontological sampling biases in fishes based on the Phanerozoic record of Great Britain. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 5–17.
- DAVIS, K. E., PISANI, D., TARVER, J. E., RUTA, M., SAKAMOTO, M., HONE, D. W. E., JENNINGS, R. and BENTON, M. J. 2008. Dinosaurs and the Cretaceous terrestrial revolution. *Proceedings of the Royal Society of London B: Biological Sciences*, **275**, 2483–2490.
- LLOYD, G. T., YOUNG, J. R. and SMITH, A. B. 2012. Taxonomic structure of the fossil record is shaped by sampling bias. *Systematic Biology*, **61**, 80–89.
- LOHR, S. 2009. *Sampling: design and analysis*, Second edition. Cengage Learning, Independence, Kentucky, 608 pp.
- MANNION, P. D., UPCHURCH, P. and CARRANO, M. T. 2011. Testing the effect of the rock record on diversity: a multidisciplinary approach to elucidating the generic richness of sauropodomorph dinosaurs through time. *Biological Reviews*, **86**, 157–181.
- MAXWELL, W. D. and BENTON, M. J. 1990. Historical tests of the absolute completeness of the fossil record of tetrapods. *Paleobiology*, **16**, 322–335.
- MCGOWAN, A. J. and SMITH, A. B. 2008. Are global Phanerozoic marine diversity curves truly global? A study of the relationship between regional rock records and global Phanerozoic marine diversity. *Paleobiology*, **34**, 80–103.
- NABOUT, J. C., ROCHA, B. D. S., CARNEIRO, F. M. and SANT'ANNA, C. L. 2013. How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. *Biodiversity and Conservation*, **22**, 2907–2918.
- NEWHAM, E., BENSON, R. and UPCHURCH, P. 2014. Mesozoic mammaliaform diversity: the effect of sampling corrections on reconstructions of evolutionary dynamics. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **412**, 32–44.
- NICOLAS, M. and RUBIDGE, B. S. 2010. Changes in Permo-Triassic terrestrial tetrapod ecological representation in the Beaufort Group (Karoo Supergroup) of South Africa. *Lethaia*, **43**, 45–59.
- NORELL, M. A. and NOVACEK, M. J. 1992. The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. *Science*, **255**, 1690–1693.
- PEARSON, M. R., BENSON, R. B. J., UPCHURCH, P., FRÖBISCH, J. and KAMMERER, C. F. 2013. Reconstructing the diversity of early terrestrial herbivorous tetrapods. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 42–49.
- PETERS, S. E. 2005. Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12326–12331.
- 2008. Environmental determinants of extinction selectivity in the fossil record. *Nature*, **454**, 626–629.
- and FOOTE, M. 2001. Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology*, **27**, 583–601.
- 2002. Determinants of extinction in the fossil record. *Nature*, **416**, 420–424.
- and HEIM, N. A. 2010. The geological completeness of paleontological sampling in North America. *Paleobiology*, **36**, 61–79.
- PURNELL, M. A. and DONOGHUE, P. 2005. Between death and data: biases in interpretation of the fossil record of conodonts. *Special Papers in Palaeontology*, **73**, 7–25.
- RAUP, D. M. 1972. Taxonomic diversity during the Phanerozoic. *Science*, **177**, 1065–1071.
- 1976. Species diversity in the Phanerozoic: an interpretation. *Paleobiology*, **2**, 289–297.
- 1977. Systematists follow the fossils. *Paleobiology*, **3**, 328–329.
- SAHNEY, S. and BENTON, M. J. 2008. Recovery from the most profound mass extinction of all time. *Proceedings of the Royal Society of London B: Biological Sciences*, **275**, 759–765.
- SEPKOSKI, J. J. Jr 1976. Species diversity in the Phanerozoic: species–area effects. *Paleobiology*, **2**, 298–303.
- 1993. Ten years in the library: new data confirm paleontological patterns. *Paleobiology*, **19**, 43–51.
- BAMBACH, R. K., RAUP, D. M. and VALENTINE, J. W. 1981. Phanerozoic marine diversity and the fossil record. *Nature*, **293**, 435–437.
- SHEEHAN, P. M. 1977. A reflection of labor by systematists? *Paleobiology*, **3**, 325–328.
- SIDOR, C. A., VILHENA, D. A., ANGIELCZYK, K. D., NESBITT, S. J., PEECOOK, B. R., STEYER, J. S., SMITH, R. M. H. and TSUJI, L. A. 2013. Provincialization of terrestrial faunas following the end-Permian mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 8129–8133.
- SMITH, A. B. 2001. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society B*, **356**, 351–367.
- 2007. Marine diversity through the Phanerozoic: problems and prospects. *Journal of the Geological Society*, **164**, 731–745.

- SMITH, R. and BOTHA-BRINK, J. 2014. Anatomy of a mass extinction: sedimentological and taphonomic evidence for drought-induced die-offs at the Permo-Triassic boundary in the main Karoo Basin, South Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **396**, 99–118.
- SMITH, A. B. and MCGOWAN, A. J. 2007. The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of western Europe? *Palaeontology*, **50**, 765–774.
- LLOYD, G. T. and MCGOWAN, A. J. 2012. Phanerozoic marine diversity: rock record modelling provides an independent test of large-scale trends. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 4489–4495.
- STANLEY, S. M. 2007. An analysis of the history of marine animal diversity. *Paleobiology*, **33**, 1–55.
- TARVER, J. E., BRADDY, S. J. and BENTON, M. J. 2007. The effects of sampling bias on Palaeozoic faunas and implications for macroevolutionary studies. *Palaeontology*, **50**, 177–184.
- DONOGHUE, P. C. J. and BENTON, M. J. 2011. Is evolutionary history repeatedly rewritten in light of new fossil discoveries? *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 599–604.
- TVERDOKHLEBOV, V. P., TVERDOKHLEBOVA, G. I., SURKOV, M. V. and BENTON, M. J. 2003. Tetrapod localities from the Triassic of the SE of European Russia. *Earth-Science Reviews*, **60**, 1–66.
- — MINIKH, A. V., SURKOV, M. V. and BENTON, M. J. 2005. Upper Permian vertebrates and their sedimentological context in the South Urals, Russia. *Earth-Science Reviews*, **69**, 27–77.
- UPCHURCH, P., MANNION, P. D., BENSON, R. B. J., BUTLER, R. J. and CARRANO, M. T. 2011. Geological and anthropogenic controls on the sampling of the terrestrial fossil record: a case study from the Dinosauria. *Geological Society, London, Special Publication*, **358**, 209–240.
- WALL, P. D., IVANY, L. C. and WILKINSON, B. H. 2009. Revisiting Raup: exploring the influence of outcrop area on diversity in light of modern sample-standardization techniques. *Paleobiology*, **35**, 146–167.
- WIGNALL, P. B. and BENTON, M. J. 1999. Lazarus taxa and fossil abundance at times of biotic crisis. *Journal of the Geological Society*, **156**, 453–456.